

LEARNING AND RISK AVERSION

A Dissertation

by

CARLOS OYARZUN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2007

Major Subject: Economics

LEARNING AND RISK AVERSION

A Dissertation

by

CARLOS OYARZUN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Rajiv Sarin
Committee Members,	Guoqiang Tian
	Brit Grosskopf
	Dante DeBlassie
Head of Department,	Amy Glass

August 2007

Major Subject: Economics

ABSTRACT

Learning and Risk Aversion. (August 2007)

Carlos Oyarzun, B.A., Universidad de Concepcion;

M.A., Universidad de Chile

Chair of Advisory Committee: Dr. Rajiv Sarin

This dissertation contains three essays on learning and risk aversion. In the first essay we consider how learning may lead to risk averse behavior. A learning rule is said to be risk averse if it is expected to add more probability to an action which provides, with certainty, the expected value of a distribution rather than when it provides a randomly drawn payoff from this distribution, for every distribution. We characterize risk averse learning rules. The result reveals that the analysis of risk averse learning is isomorphic to that of risk averse expected utility maximizers. A learning rule is said to be monotonically risk averse if it is expected to increase the probability of choosing the actions whose distribution second-order stochastically dominates all others in every environment. We characterize monotonically risk averse learning rules.

In the second essay we analyze risk attitudes for learning within the mean-variance paradigm. A learning rule is variance-averse if the expected reduced distribution of payoffs in the next period has a smaller variance than that of the current reduced distribution, in every set where all the actions provide the same expected payoff. A learning rule is monotonically variance-averse if it is expected to add probability to the set of actions that have the smallest variance in the set, when all the actions have the same expected payoff. A learning rule is monotonically mean-variance-averse if it is expected to add probability to the set of actions that have the

highest expected payoff and smallest variance whenever this set is not empty. We characterize monotonically variance-averse and monotonically mean-variance-averse learning rules.

In the last essay we analyze the social learning process of a group of individuals. We say that a learning rule is first-order monotone if the number of individuals that play actions with first-order stochastic dominant payoff distributions is expected to increase. We characterize these learning rules.

ACKNOWLEDGMENTS

I am deeply grateful for guidance and support from my committee chair, Dr. Rajiv Sarin, and my committee members, Drs. Guoqiang Tian, Brit Grosskopf, and Dante DeBlassie. I would like to thank each of my professors at Texas A&M University, especially Drs. William Neilson and Timothy Gronberg. I thank Johannes Ruf for joint work on Chapter IV of this dissertation. I thankfully acknowledge financial support from the Department of Economics, the Private Enterprise Research Center (PERC) at Texas A&M University and the Bradley Foundation. I also thank Tyffanne Rowan for administrative assistances. I would like to thank my parents for encouraging me to pursue higher education. Thanks to my friends in College Station; especially Paan Jindapon and Diego Escobari. Finally, I would like to thank Maria Teresa Erpelding, my first calculus-analysis teacher, and all my professors of mathematics for trying to teach me to think with discipline and rigor.

TABLE OF CONTENTS

CHAPTER		Page
I	INTRODUCTION	1
II	LEARNING AND RISK AVERSION	5
	A. Introduction	5
	B. Framework	8
	C. Risk-Averse Learning	11
	D. Monotonic Risk Aversion	14
	E. Examples	21
	F. Discussion	23
III	MEAN AND VARIANCE RESPONSIVE LEARNING	26
	A. Introduction	26
	B. Framework	30
	C. Weak Unbiasedness	33
	D. Learning and Behavior Toward Variance	40
	E. Monotonically Mean-Variance-Averse Rules	44
	F. Conclusion	47
IV	MONOTONE IMITATION	49
	A. Introduction	49
	B. Framework	55
	C. First-Order Monotone Imitation Rules	58
	D. Large Population Dynamics	68
	E. Examples	70
	F. Individual Monotonicity	73
	G. Second-Order Monotonicity	76
	H. Discussion	80
V	SUMMARY	82
	REFERENCES	84
	VITA	88

CHAPTER I

INTRODUCTION

One of the advantages of Expected Utility Theory is that it provides a precise notion of risk aversion. Furthermore, the way to specify risk-averse preferences is well known; namely, they are described by the concavity of the Bernoulli utility functions. In applications, such as the construction of general equilibrium models for the analysis of asset prices, the risk attitudes of the decision maker (and their intensity) are fundamental for the analysis of the predictions of the model. However, perhaps, one may not feel comfortable with the expected utility paradigm. This could happen because of a number of reasons; for example it may be the case that the set up suggests that the assumptions in which this theory lies on are too strong. Or, it could be that the implications of the model contradict some fundamental features of the data under study. In either case attention is directed to alternative theories of decision making. In applications, we often find economic agents who make decisions according to alternative paradigms. Unfortunately, the attitudes to risk of these agents seldom are known. Many times, the proper notion of risk aversion and, consequently, the way to specify it are missing.

The notion of risk aversion arises naturally in the analysis of preferences; one person may prefer a safe lottery to a risky lottery. Nevertheless, risk aversion may arise in different contexts as well; for example, people may tend to choose safer actions over time. It may reflect a systematic attitude in search of decisions involving less uncertainty. This dissertation tries to formalize this type of risk aversion.

The objective of this dissertation is to provide suitable notions of risk attitudes

This dissertation follows the style of Journal of Economic Theory.

for learning and to obtain characterizations specifying the conditions that must be satisfied for the learning rules that verify those properties. The analysis reveals strong connections between our theory of risk aversion for learning and Expected Utility Theory. Many of the definitions and results in our theory rely on the mathematical structures derived for the analysis of Expected Utility Theory.

The first essay of this dissertation, Chapter II, considers an adaptive learner who repeatedly faces a problem in which she must choose one action out of a finite set of actions. Our model is a probabilistic choice model in the sense that each action is chosen with some probability. After one action is chosen, the decision maker observes the payoff she obtained. This payoff is a random variable whose probability distribution is unknown to the decision maker. After choosing an action and obtaining the corresponding payoff, the decision maker updates her behavior by changing the probability with which she will choose each action in the next period. We formalize this process using the learning rule concept. A learning rule is a function mapping the chosen action and the obtained payoff to a vector of probabilities of choosing each action in the next period. We introduce two notions of risk aversion for learning rules. These notions are based on the expected change in the probability of choosing each action in the set; and risk is measured using Rothschild and Stiglitz [31] second-order stochastic dominance (sosd) concept. We say that a learning rule is risk-averse if for every distribution the action is expected to increase the probability of being chosen when it pays the expected value of the distribution with probability one more than when it provides a random payoff drawn from the distribution itself. We say that a learning rule is monotonically risk-averse if it is expected to add probability to the set of actions that second-order stochastically dominate all the other actions in the set.

In the second essay of this dissertation we consider a decision maker facing the

same problem described above. Our objective is, again, to propose notions of risk aversion for learning rules; however, in this chapter, we use the variance of a distribution to measure risk and, accordingly, our definitions of risk aversion are based on this concept. We say that a learning rule is variance-averse if the expected reduced-distribution of payoffs in the next period has a smaller variance than that of the current reduced-distribution, in every set where all the actions provide the same expected payoff. We also call a learning rule monotonically variance-averse if it is expected to add probability to the set of actions that have the smallest variance, when all the actions have the same expected payoff. Finally, we call a learning rule monotonically mean-variance-averse if it is expected to add probability to the subset of actions that have the highest expected payoff and smallest variance whenever this subset is not empty.

Our definition of risk-averse learning rules is inspired by the definition of risk aversion in Expected Utility Theory. Our notion of monotonically risk-averse learning rules, instead, is related to the work of Börgers, Morales and Sarin [4]. They study monotone learning rules. A learning rule is monotone if in every environment it is expected to add probability mass to the set of expected payoff maximizing actions. Therefore our notion of monotonicity replaces the expected value of the payoff in Börgers et al. with the expected value of an arbitrary concave function. Our notion of monotonically variance-averse learning rules relate to Börgers et al.'s notion in the same way, but it replaces expected payoff with variance, and our learning rules seek to avoid it. Our notion of monotone mean-variance aversion combines monotonicity in Börgers et al. with monotone variance aversion.

In the third essay of this dissertation we study the dynamics of choices made by a population of individuals. In this model each individual has to choose one action every period. The set of available actions and the corresponding distributions of payoffs

remain the same from one period to the next. While the set of actions is known, their payoff distributions are not. Each individual is able to recall the action she played in the last period and the payoff she obtained; furthermore, she recalls the action chosen by another member of the population and the payoff he got. We analyze the imitation rule concept. An imitation rule of a given individual is a function mapping both the played and observed actions, and the corresponding obtained payoffs, to a vector of probabilities with which she will play each action in the next period. We call an imitation rule first-order monotone if the number of individuals that play the actions more likely to provide higher payoffs is expected to increase in every period. We characterize these rules and also provide a sufficient condition under which all the individuals in the population converge to choose these actions. We provide the analysis of two extensions. The first extension studies properties defined in terms of the performance at the individual level. The second extension explains how to introduce concerns for risk within the analysis provided in this essay.

The analytical framework of the third essay is inspired by Schlag's [33] imitation model; however, our notion of first-order monotone rules is more similar to the concept of monotonicity of Börgers et al. than to the concepts studied by Schlag.

CHAPTER II

LEARNING AND RISK AVERSION

A. Introduction

Expected Utility Theory (EUT) explains behavior in terms of preferences over, possibly subjective, probability distributions. Preferences are taken as innate and exogenously given. In this essay we take an alternate approach in which the decision maker need not know that she is choosing among probability distributions. The behavior of the agent is explained by the manner in which the agent responds to experience. It is the response to experience that our approach takes as exogenously given. This essay reveals that the analysis of risk aversion can be developed, in this alternate approach, in much the same manner as in EUT.

We describe the behavior of an individual by the probabilities with which she chooses her alternative actions.¹ The manner in which the agent responds to experience is described by a *learning rule*. For given behavior today, a learning rule is a mapping from the action played in this period and the payoff obtained to behavior in the next period. We refer to the payoff distribution associated with each action as the *environment*. The environment is assumed to be unknown to the decision maker.

The focus of this essay is on how learning through experience explains behavior towards risk. In EUT, preferences are said to be risk-averse if, for every probability distribution, the expected value of the distribution is preferred to the distribution itself. In order to define risk aversion for a learning rule, we consider two environ-

¹This is commonly assumed in models of learning. For a recent axiomatic foundation of probabilistic behavior in decision making see Gul and Pesendorfer [18].

ments that differ only in the distributions associated with one of the actions. In the first environment, the action provides a payoff drawn randomly from a given distribution, whereas in the second environment the expected payoff of that distribution is provided with certainty. We say that a learning rule is *risk-averse* if, for any distribution, the learning rule is expected to add more probability to the action in the second environment than in the first. Furthermore, we require the above to hold regardless of the distributions over payoffs obtained from the other actions. Formally, the definition of when a learning rule is risk-averse replaces the greater expected utility of the action that gives the expected value (rather than the distribution itself), that describes a risk-averse expected utility maximizer, with being expected to add greater probability mass to the action that gives the expected value (rather than the distribution itself).

Our first result shows that a learning rule is risk-averse if and only if the manner in which it updates the probability of the chosen action is a concave function of the payoff it receives. This result allows us to develop a theory of the risk attitudes of learning rules that is isomorphic to that of the risk attitudes of expected utility maximizers. Our analysis, hence, provides a bridge between decision theory and learning theory.²

In contrast to decision theory in which the agent selects an (optimal) action based on her beliefs about the distributions over payoffs of each action, a learning rule specifies how probability is moved among actions. Therefore, we could ask if the manner in which the learning rule shifts probability between actions results in the optimal action(s) being chosen with increased probability. This motivates us to investigate *monotonically risk-averse* learning rules, which require that the learning

²For earlier attempts to relate decision theory and learning theory see March [22] and Simon [35].

rule is expected to add probability mass on the set of actions whose distributions second-order stochastically dominate those of all other actions, in every environment.

We provide a characterization of monotonically risk-averse learning rules. The result shows that how the learning rule updates the probability of unchosen actions, in response to the payoff obtained from the chosen action, plays a critical role. This response has to be a convex function of the payoff for each unchosen action. Furthermore, every monotonically risk-averse learning rule satisfies a property we call *impartiality*. We say that a learning rule is impartial if there is no expected change in the probability of playing each action whenever all actions have the same payoff distribution. The restrictions imposed by impartiality on the functional form of the learning rule, together with the convex response of unchosen actions, imply that every monotonically risk-averse learning rule is risk-averse.

We also characterize *first-order monotone* learning rules which require that the learning rule be expected to increase probability mass on the set of actions whose distributions first order stochastically dominate all others. We show that a learning rule is first-order monotone if and only if it is impartial and the learning rule updates the probability of each unchosen action according to a decreasing function of the payoff obtained from the chosen action. This latter condition provides the analogue of the requirement that Bernoulli utility functions are increasing in EUT. This essay, therefore, provides a classification of learning rules that is much like the manner in which EUT classifies Bernoulli utility functions. In particular, our results allow us to determine how any given learning rule responds to any distribution over payoffs.

Our second and third characterizations provide a generalization of the results obtained by Börgers et al., who consider learning rules that are expected to increase probability on the expected payoff maximizing actions in every environment. They call such rules monotone. It is straightforward to show that monotone learning rules

are monotonically risk-averse and first-order monotone. In Section E of this essay, we show by example, that the class of monotonically risk-averse and first-order monotone learning rules includes several well known learning rules that are not monotone.

There are some papers that investigate how learning respond to risk. March [22] and Burgos [6] investigate specific learning rules by way of simulations. Both consider an environment in which the decision maker has two actions, one of which gives the expected value of the other (risky) action with certainty. As in our analysis, the learning rules they consider update behavior using only the information on the payoff obtained from the chosen action. For the specific rules they consider, they show that they all choose the safe action more frequently over time. Denrell [12] analytically shows that a class of learning rules choose the safe action more frequently in the long run. His result is obtained even if the decision maker follows an optimal policy of experimentation.³

B. Framework

Let A be the finite set of actions available to the decision maker. Action $a \in A$ gives payoffs according to the distribution function F_a . We shall refer to $F = (F_a)_{a \in A}$ as the environment the individual faces and we assume that it does not change from one period to the next. The agent knows the set of actions A but not the distributions F . The decision maker is assumed to know the finite upper and lower bounds on the set of possible payoffs $X = [x_{\min}, x_{\max}]$. We may think of payoffs as monetary magnitudes.

The behavior of the individual is described by the probability with which she

³For a discussion of the manner in which evolution may affect behavior toward risk see Dekel and Schotchmer [10] and Robson [27], [28].

chooses each action. Let behavior today be given by the mixed action vector $\sigma \in \Delta(A)$, where $\Delta(A)$ denotes the set of all probability distributions over A . We assume that there is a strictly positive probability that each action is chosen today.⁴ Taking the behavior of the agent today as given, a learning rule L specifies her behavior tomorrow given the action $a \in A$ she chooses and the monetary payoff $x \in X$ she obtains today. Hence, $L : A \times X \rightarrow \Delta(A)$. The learning rule should be interpreted as a “reduced form” of the true learning rule. The true learning rule may, for example, specify how the decision maker updates her beliefs about the payoff distributions in response to her observations and how these beliefs are translated into behavior. If one combines the two steps of belief adjustment and behavior change we get a learning rule as we define.⁵

Let $L_{(a',x)}(a)$ denote the probability with which a is chosen in the next period if a' was chosen today and a payoff of x was received. For a given learning rule L and environment F , the expected movement of probability mass on action a is $f(a) := \sum_{a' \in A} \sigma_{a'} \int L_{(a',x)}(a) dF_{a'}(x) - \sigma_a$.

Denote the expected payoff associated with F_a by π_a . Let the distributions over payoffs of actions other than a be denoted by F_{-a} . The definition of when a learning rule is risk-averse requires that if we replace the distribution F_a with another distribution \tilde{F}_a which puts all probability on π_a and keep F_{-a} fixed, then the learning rule is expected to add more probability mass to a when it gives payoffs according to \tilde{F}_a than when it gives payoffs according to F_a . This should be true for all a , F_a and F_{-a} . More formally, we introduce a second associated environment $\tilde{F} = (\tilde{F}_a, \tilde{F}_{-a})$ in

⁴This assumption can be dropped with minor changes in the proofs though it would require additional notation.

⁵Notice that we do not restrict the state space of the learning rule to be the probability simplex $\Delta(A)$. The examples in Section E illustrate this.

which $\tilde{F}_{-a} = F_{-a}$. Hence, environment \tilde{F} has the same set of actions as environment F and the distribution over payoffs of all actions other than a are as in F . Let $\tilde{f}(a)$ denote the expected movement of probability mass on action a in the associated environment \tilde{F} .

Definition 1 *A learning rule L is risk-averse if for all a and F , if \tilde{F}_a places all probability mass on π_a then $\tilde{f}(a) \geq f(a)$.*

Risk seeking and risk neutral learning rules may be defined in the obvious manner. As the analysis of such learning rules involves a straightforward extension we do not pursue it further in the sequel. The risk aversion of the learning rule may be considered a “local” concept as we have taken the current state as given. If the current state of learning is represented by an element $s \in S$ which is a subset of finite dimensional Euclidean space then the risk aversion of a learning rule “in the large” could be considered by defining a learning rule to be globally risk-averse if it is risk-averse at all states $s \in S$. Our work provides a first step in the analysis of globally risk-averse learning rules.

The contrast between when a learning rule is risk-averse and when an expected utility maximizer is risk-averse is instructive. In EUT an individual is called risk-averse if for all distributions F_a the individual prefers \tilde{F}_a to F_a . Hence, the von Neumann-Morgenstern utilities v satisfy

$$v(\tilde{F}_a) = \int u(x) d\tilde{F}_a(x) \geq \int u(x) dF_a(x) = v(F_a), \quad (2.1)$$

where $u(\cdot)$ is often referred to as the Bernoulli utility function. A learning rule is called risk-averse if for all actions a and distributions F_a the learning rule is expected to add more probability mass to an action that gives π_a with certainty than to an action that gives payoffs according to F_a regardless of the distributions F_{-a} . Hence,

risk aversion in learning requires that $\tilde{f}(a) \geq f(a)$, i.e.,

$$\sum_{a' \in A} \sigma_{a'} \int L_{(a',x)}(a) d\tilde{F}_{a'}(x) - \sigma_a \geq \sum_{a' \in A} \sigma_{a'} \int L_{(a',x)}(a) dF_{a'}(x) - \sigma_a. \quad (2.2)$$

Notice that, whereas $v(\cdot)$ in EUT depends only on the payoff distribution of a single action, $f(\cdot)$ in learning theory depends on the distribution of the entire vector of distributions.

C. Risk-Averse Learning

In this section we state our results regarding risk-averse learning rules and their relationship to results concerning risk-averse expected utility maximizers. The following definition provides some useful terminology.

Definition 2 *A learning rule L is own-concave if for all a , $L_{(a,x)}(a)$ is a concave function of x .*

A learning rule tells us how the probability of *each* action $a' \in A$ is updated upon choosing any action a and receiving a payoff x . Own-concavity of a learning rule places a restriction only on the manner in which the updated probability of action a depends upon x given that a is chosen.

Proposition 1 *A learning rule L is risk-averse if and only if it is own-concave.*

Proof. We begin by proving that every own-concave learning rule is risk-averse. Consider any own-concave learning rule L and environment $F = (F_a, F_{-a})$. Construct the associated environment \tilde{F} in which \tilde{F}_a places all probability mass on π_a (and $F_{-a} = \tilde{F}_{-a}$). By Jensen's inequality,

$$L_{(a,\pi_a)}(a) \geq \int L_{(a,x)}(a) dF_a(x) \quad (2.3)$$

\Longleftrightarrow

$$\int L_{(a,x)}(a) d\tilde{F}_a(x) \geq \int L_{(a,x)}(a) dF_a(x) \quad (2.4)$$

 \Longleftrightarrow

$$\sigma_a \int L_{(a,x)}(a) d\tilde{F}_a(x) + \sum_{a' \neq a} \sigma_{a'} \int L_{(a',x)}(a) d\tilde{F}_{a'}(x) - \sigma_a \quad (2.5)$$

$$\geq \sigma_a \int L_{(a,x)}(a) dF_a(x) + \sum_{a' \neq a} \sigma_{a'} \int L_{(a',x)}(a) dF_{a'}(x) - \sigma_a \quad (2.6)$$

 \Longleftrightarrow

$$\tilde{f}(a) \geq f(a). \quad (2.7)$$

Hence, the learning rule is risk-averse.

We now turn to prove that every risk-averse learning rule L is own-concave. We argue by contradiction. Suppose L is risk-averse but not own-concave. Because L is not own-concave there exists an action a , payoffs $x', x'' \in [x_{\min}, x_{\max}]$ and $\lambda \in (0, 1)$ such that

$$L_{(a, \lambda x' + (1-\lambda)x'')}(a) < \lambda L_{(a, x')}(a) + (1-\lambda) L_{(a, x'')}(a). \quad (2.8)$$

Now consider an environment F in which F_a gives x' with probability λ and x'' with probability $(1-\lambda)$ and the distributions of the other actions are given by F_{-a} . Consider the associated environment \tilde{F} in which \tilde{F}_a gives $\pi_a = \lambda x' + (1-\lambda)x''$ with probability one. Hence,

$$\int L_{(a,x)}(a) d\tilde{F}_a(x) = L_{(a, \pi_a)}(a) \quad (2.9)$$

$$< \lambda L_{(a, x')}(a) + (1-\lambda) L_{(a, x'')}(a) \quad (2.10)$$

$$= \int L_{(a,x)}(a) dF_a(x). \quad (2.11)$$

which implies $\tilde{f}(a) < f(a)$ by the argument above. Hence, the rule is not risk-averse as we had assumed and we obtain a contradiction. \square

Proposition 1 shows that the own-concavity of a learning rule in learning theory plays an analogous role as the concavity of the Bernoulli utility function in EUT. In the latter theory the curvature properties of a Bernoulli utility function explain the individuals attitudes towards risk. In the theory of learning, the manner in which the learning rule updates the probability of the chosen action in response to the payoff it obtains explains how learning responds to risk. The proof reveals that for any action a the distributions of actions $a' \neq a$ do not play any role when we compare $\tilde{f}(a)$ and $f(a)$. This has the consequence that the theory of risk-averse learning rules is isomorphic to the theory of risk-averse expected utility maximizers.

For example, if $L_{(a,x)}(a)$ is a twice differentiable function of x , we can adapt the well known Arrow-Pratt measure of absolute risk aversion (Arrow [2], Pratt [26]) to find an easy measure of the risk aversion of a learning rule. Specifically, we define the coefficient of absolute risk aversion of a learning rule L for action a as

$$ar_{L_a}(x) = -\frac{\partial^2 L_{(a,x)}(a) / \partial x^2}{\partial L_{(a,x)}(a) / \partial x}. \quad (2.12)$$

In EUT a distribution F_a is said to be more risky than another \tilde{F}_a if both have the same mean and every risk-averse person prefers \tilde{F}_a to F_a (see, e.g., Rothschild and Stiglitz [31]). In this case it is usually said that \tilde{F}_a second-order stochastically dominates (sosd) F_a . The following Corollary shows that an analogous result applies in our case.

Corollary 1 *\tilde{F}_a second-order stochastically dominates F_a if and only if $\tilde{f}(a) \geq f(a)$ for all a for every risk-averse learning rule.*

Proof. $\tilde{f}(a) \geq f(a)$ for every risk-averse learning rule

\Longleftrightarrow

$$\int L_{(a,x)}(a) d\tilde{F}_a(x) \geq \int L_{(a,x)}(a) dF_a(x) \text{ for every own-concave } L$$

\Longleftrightarrow

\tilde{F}_a second-order stochastically dominates F_a . \square

For risk-averse learning rules, imposing the requirement that probabilities of all actions must sum to one provides the obvious restrictions when there are only two actions. However, few restrictions are imposed when there are three or more actions. The property we study in the next section provides such restrictions.

D. Monotonic Risk Aversion

The definition of a risk-averse learning rule was inspired by standard decision theory. Learning, however, differs in many respects from choice. Whereas in decision theory a single (and optimal) action is chosen, in learning theory probability is moved between actions. For learning, it then appears reasonable to ask whether probability on the optimal action is increased from one period to the next. Our next definition introduces such a property. Specifically, a learning rule is said to be monotonically risk-averse if it is expected to increase probability on the best actions, in a sosd sense, in every environment.

Let A^* denote the set of actions that second-order stochastically dominate all other actions. That is, $A^* := \{a \in A : F_a \text{ sosd } F_{a'} \forall a' \in A\}$. Clearly, if $A^* = A$ we have that $F_a = F_{a'}$ for all $a, a' \in A$. For any subset $\hat{A} \subset A$, let $f(\hat{A}) := \sum_{a \in \hat{A}} f(a)$.

Definition 3 *A learning rule L is monotonically risk-averse if in all environments we have that $f(A^*) \geq 0$.*

Correspondingly, we say that a learning rule is monotonically risk seeking if $f(A^*) \leq 0$ in every environment and a learning rule is monotonically risk neutral if it is monotonically risk-averse and monotonically risk seeking. The analysis of such rules is analogous to the analysis of monotonically risk-averse learning rules provided

below. Note that, when A^* is empty, expected utility maximization by a risk-averse agent places no specific restrictions on behavior. This has the analogue, in our analysis, that no restrictions are placed on the movement of probability when A^* is empty.

Now, we proceed to characterize monotonically risk-averse learning rules. The following definition introduces some useful terminology.

Definition 4 *A learning rule L is cross-convex if for all a , $L_{(a',x)}(a)$ is convex in x for all $a' \neq a$.*

We shall see in the next result that all monotonically risk-averse learning rules have the feature that if all actions have the same distribution of payoffs then there is no expected movement in probability mass on any action. We call such learning rules *impartial*.

Definition 5 *A learning rule L is impartial if $f(a) = 0$ for all a whenever $F_a = F_{a'}$ for all $a, a' \in A$.*

The set of impartial learning rules is related to the unbiased learning rules studied in Börgers et al. Unbiasedness requires that no probability mass is expected to be moved among actions when all have the same expected payoff. Clearly, the set of impartial learning rules is larger than the set of unbiased learning rules. Furthermore, it is straightforward to see that unbiased learning rules cannot respond to aspects of the distribution of payoffs other than the mean.⁶

Proposition 2 *A learning rule L is monotonically risk-averse if and only if (i) $\sigma_a = \sum_{a' \in A} \sigma_{a'} L_{(a',x)}(a)$ for all a , and (ii) L is cross-convex.*

⁶From the analysis below, it is not difficult to see that a learning rule is unbiased if and only if it is monotonically risk neutral.

Our proof begins with two Lemmas. The first shows that all monotonically risk-averse learning rules are impartial and the second characterizes impartial learning rules.

Lemma 1 *If the learning rule L is monotonically risk-averse then it is impartial.*

Proof. The proof is by contradiction. Suppose L is monotonically risk-averse but there exists an environment F with $A = A^*$ and $f(a) > 0$ for some $a \in A$. If F_a does not place strictly positive probability on (x_{\min}, x_{\max}) , then consider the environment \hat{F} such that, for all action $a \in A$, the probabilities of x_{\min} and x_{\max} are $(1 - \varepsilon)$ times their corresponding probabilities in the environment F , and the probability of some $x \in (x_{\min}, x_{\max})$ is ε . If F_a places strictly positive probability on (x_{\min}, x_{\max}) , then let $\hat{F} = F$. We now construct the environment \tilde{F} in which \tilde{F}_a is a mean preserving spread of \hat{F}_a and $\tilde{F}_{a'} = \hat{F}_{a'}$ for all $a' \neq a$. Specifically, suppose that \tilde{F}_a is obtained by assigning to every interval $I \subset [x_{\min}, x_{\max}]$ only $(1 - \varepsilon)$ times the probability it had under \hat{F}_a and then adding $(\hat{\pi}_a - x_{\min})\varepsilon/(x_{\max} - x_{\min})$ on the probability of x_{\max} and $(x_{\max} - \hat{\pi}_a)\varepsilon/(x_{\max} - x_{\min})$ on the probability of x_{\min} . By construction, $\tilde{F}_{a'} \succ \tilde{F}_a$ for all $a' \neq a$. It follows that $\tilde{A}^* = A \setminus \{a\}$. Since $\tilde{f}(a)$ can be written as a continuous function in ε , there exists a small enough ε such that $\tilde{f}(a) > 0$. This contradicts that L is monotonically risk-averse. \square

Lemma 2 *A learning rule L is impartial if and only if for all $a \in A$ and $x \in X$, it satisfies $\sigma_a = \sum_{a' \in A} \sigma_{a'} L_{(a', x)}(a)$.*

Proof. *Necessity.*

Consider an environment where all the actions pay x with probability one. Then, for all $a \in A$,

$$f(a) = \sum_{a' \in A} \sigma_{a'} L_{(a', x)}(a) - \sigma_a. \quad (2.13)$$

Therefore, in order to be impartial L must satisfy

$$\sigma_a = \sum_{a' \in A} \sigma_{a'} L_{(a',x)}(a). \quad (2.14)$$

Sufficiency.

Consider the environment F such that $F_a = F_{a'}$ for all $a, a' \in A$.

$$f(a) = \sum_{a' \in A} \sigma_{a'} \int L_{(a',x)}(a) dF_{a'}(x) - \sigma_a \quad (2.15)$$

$$= \int \sum_{a' \in A} \sigma_{a'} L_{(a',x)}(a) dF_a(x) - \sigma_a \quad (2.16)$$

$$= 0. \quad (2.17)$$

The second statement follows from the fact that all the distributions are the same, and the third statement follows from the hypothesis. \square

Proof of Proposition 2.

Sufficiency.

Consider $a \in A^*$

$$f(a) = \sigma_a \int L_{(a,x)}(a) dF_a(x) + \sum_{a' \neq a} \sigma_{a'} \int L_{(a',x)}(a) dF_{a'}(x) - \sigma_a \quad (2.18)$$

$$= \sum_{a' \neq a} \sigma_{a'} \left[\int L_{(a',x)}(a) dF_{a'}(x) - \int L_{(a',x)}(a) dF_a(x) \right] \quad (2.19)$$

$$\geq 0. \quad (2.20)$$

The second statement follows from Lemmas 1 and 2 and the last inequality follows from the fact that $a \in A^*$ and the convexity of the functions $L_{(a',x)}(a)$ for all $a' \in A \setminus \{a\}$.

Necessity.

We argue by contradiction. Suppose that for some $a \in A$ and some $a' \in A \setminus \{a\}$, $L_{(a',x)}(a)$ is not convex. Therefore there exists $x', x'', \lambda \in (0, 1)$ and $x := \lambda x' + (1 -$

$\lambda)x''$ such that $\lambda L_{(a',x')}(a) + (1 - \lambda)L_{(a',x'')}(a) < L_{(a',x)}(a)$. Consider an environment where $a' \in A \setminus \{a\}$ pays x' with probability λ , and x'' with probability $(1 - \lambda)$. Action a pays x with probability one, and all the other actions in the set, if any, pay x with probability $1 - \varepsilon$, x' with probability $\varepsilon\lambda$, and x'' with probability $\varepsilon(1 - \lambda)$. Clearly, $A^* = \{a\}$. From the sufficiency part we know

$$f(a) = \sum_{a' \neq a} \sigma_{a'} \left[\int L_{(a',x)}(a) dF_{a'}(x) - \int L_{(a',x)}(a) dF_a(x) \right] \quad (2.21)$$

$$= \sigma_{a'} [\lambda L_{(a',x')}(a) + (1 - \lambda)L_{(a',x'')}(a) - L_{(a',x)}(a)] \quad (2.22)$$

$$+ \varepsilon \sum_{a'' \neq a, a'} \sigma_{a''} [\lambda L_{(a'',x')}(a) + (1 - \lambda)L_{(a'',x'')}(a) - L_{(a'',x)}(a)]. \quad (2.23)$$

Therefore, for small enough ε , $f(a) < 0$. \square

Monotonic risk aversion places restrictions on how the learning rule updates the probability of each unchosen action as a function of the action chosen and payoff obtained. In particular, it requires this function be convex in the payoff received, for each unchosen action. Furthermore, we show that all such rules have the weak consistency property of impartiality. Because every impartial and cross-convex learning rule is own-concave we have the following Corollary.

Corollary 2 *Every monotonically risk-averse learning rule L is risk-averse.*

The analysis above can be extended to the concept of first-order stochastic dominance. In EUT a distribution F_a is said to first order stochastically dominate (fod) another \tilde{F}_a if every individual with an increasing (Bernoulli) utility function prefers the former. In the context of our analysis, we would like to identify the learning rules that are expected to add probability mass on the set of actions whose distributions fod the distributions of all the other actions, in every environment. We call such learning rules *first-order monotone*. Let $A^{**} := \{a \in A : a \text{ fod } a' \text{ for all } a' \in A\}$.

Definition 6 *A learning rule L is first-order monotone if $f(A^{**}) \geq 0$ in every environment.*

First-order monotone learning rules can be characterized in the same manner as monotonically risk-averse learning rules. In particular, these rules need to be impartial but instead of being cross-convex they require the response of the probabilities of playing the unchosen actions to be decreasing in the obtained payoff.

Definition 7 *A learning rule L is cross-decreasing if for all a , $L_{(a',x)}(a)$ is decreasing in x for all $a' \neq a$.*

Proposition 3 *A learning rule is first-order monotone if and only if it satisfies (i) $\sigma_a = \sum_{a' \in A} \sigma_{a'} L_{(a',x)}(a)$ for all $a \in A$ and (ii) L is cross-decreasing.*

In order to prove this result we begin with the following Lemma, the proof of which closely parallels that of Lemma 1.

Lemma 3 *Every first-order monotone learning rule L is impartial.*

Proof. We argue by contradiction. Consider an environment F with $A = A^{**}$ and suppose that $f(a) < 0$ for some $a \in A$. Now we construct an environment \hat{F} where $\hat{A}^{**} = \{a\}$. We construct \hat{F}_a by assigning to every interval $I \subset X$ only $(1 - \varepsilon)$ times the probability it had under F_a and adding ε to the probability of x_{\max} . We construct $\hat{F}_{a'}$ for all $a' \in A \setminus \{a\}$ by assigning to every interval $I \subset X$ only $(1 - \varepsilon)$ times the probability it had under $F_{a'}$ and then adding ε to the probability of x_{\min} . Clearly, $\hat{A}^{**} = \{a\}$. Since $\hat{f}(a)$ can be written as a continuous function in ε , for small enough ε we have $\hat{f}(a) < 0$. Therefore L is not first-order monotone. \square

Proof of Proposition 3.

Necessity.

The necessity of (i) follows from Lemma 3 and Lemma 2. To prove the necessity of (ii) we argue by contradiction. Suppose that for some $a \in A$ and $a' \in A \setminus \{a\}$, there are x and x' with $x' < x$ and $L_{(a',x)}(a) > L_{(a',x')}(a)$. Consider the environment F where action a pays x with probability one and action a' pays x' with probability one. All the other actions $a'' \in A \setminus \{a, a'\}$, if any, pay x with probability $1 - \varepsilon$ and x' with probability ε . Clearly, $A^{**} = \{a\}$. From Lemma 2 and Lemma 3, we have that

$$f(a) = \sum_{a' \neq a} \sigma_{a'} \left[\int L_{(a',x)}(a) dF_{a'}(x) - \int L_{(a',x)}(a) dF_a(x) \right] \quad (2.24)$$

$$= \sigma_{a'} [L_{(a',x')}(a) - L_{(a',x)}(a)] + \sum_{a'' \neq a, a'} \varepsilon \sigma_{a''} [L_{(a'',x')}(a) - L_{(a'',x)}(a)] \quad (2.25)$$

For small enough ε , $f(a) < 0$, which contradicts first-order monotonicity.

Sufficiency.

As in the proof of Proposition 2, consider $a \in A^{**}$, then

$$f(a) = \sum_{a' \neq a} \sigma_{a'} \left[\int L_{(a',x)}(a) dF_{a'}(x) - \int L_{(a',x)}(a) dF_a(x) \right] \quad (2.26)$$

$$\geq 0. \quad (2.27)$$

The last inequality follows from the fact that $a \in A^{**}$ and the fact that $L_{(a',x)}(a)$ is decreasing for all $a' \in A \setminus \{a\}$. \square

The notion of risk-averse learning of Section C may be extended to first-order stochastic dominance in a similar manner. We can identify a set of learning rules such that for every action a and distribution F_a , if that distribution is replaced by a distribution \tilde{F}_a , such that \tilde{F}_a fofd F_a , then $\tilde{f}(a) \geq f(a)$, in every environment F . It is easy to show that this set of learning rules is equivalent to the set of learning rules for which $L_{(a,x)}(a)$ is increasing in x for all $a \in A$.

E. Examples

The Cross learning rule (Cross [8]) is given by

$$L_{(a,x)}(a) = \sigma_a + (1 - \sigma_a)x \quad (2.28)$$

$$L_{(a',x)}(a) = \sigma_a - \sigma_a x \quad \forall a' \neq a, \quad (2.29)$$

where $x \in [0, 1]$. It is easily seen that the Cross learning rule is impartial. Furthermore, its cross-components are affine transformations of x . Therefore this rule is monotonically risk neutral and hence it is also risk neutral. It is also clearly first-order monotone.

The Roth and Erev [29] learning rule describes the state of learning s of an agent by a vector $v \in R_{++}^{|A|}$. The vector v describes the decision makers “attraction” to choose any of her $|A|$ actions. Given v , the agents behavior is given by $\sigma_a = v_a / \sum_{a'} v_{a'}$ for all a . If the agent plays a and receives a payoff of x then she adds x to her attraction to play a , leaving all other attractions unchanged. Hence, the Roth and Erev learning rule is given by

$$L_{(a,x)}^v(a) = \frac{v_a + x}{\sum_{a''} v_{a''} + x} \quad (2.30)$$

$$L_{(a',x)}^v(a) = \frac{v_a}{\sum_{a''} v_{a''} + x} \quad \forall a' \neq a, \quad (2.31)$$

where $x \in [0, x_{\max}]$ and the superscript v on the learning rule defines it at that state of learning. Using Lemma 2, it is easy to check that this rule is impartial. Observing that the cross-components $L_{(a',x)}^v(a)$ are decreasing convex functions of x for all $a' \neq a$, we see that this learning rule is first-order monotone and monotonically risk-averse. The coefficient of absolute risk aversion of this learning rule is $ar_{L_a} = 2 / (\sum_{a'} v_{a'} + x)$ for all a . Clearly, ar_{L_a} decreases as $\sum_{a'} v_{a'}$ increases and hence this learning rule exhibits declining absolute risk aversion. Note that this rule satisfies none of the

properties studied by Börgers et al. who have shown that this rule is neither monotone nor unbiased.

Our next example considers the weighted return model studied in March [22]. This learning rule is risk-averse but may not be monotonically risk-averse. The state of learning is described by a vector of attractions $v \in R_{++}^{|A|}$. Given v , the agents behavior is given by $\sigma_a = v_a / \sum_{a' \in A} v_{a'}$ for all a . If action a is chosen and receives a payoff of x then she adds $\beta(x - v_a)$ to her attraction of a , where $\beta \in (0, 1)$ is a parameter, leaving all other attractions unchanged. Thus, the learning rule may be written as

$$L_{(a,x)}^v(a) = \frac{v_a + \beta(x - v_a)}{\sum_{a'' \in A} v_{a''} + \beta(x - v_a)} \quad (2.32)$$

$$L_{(a',x)}^v(a) = \frac{v_a}{\sum_{a'' \in A} v_{a''} + \beta(x - v_{a'})} \quad \forall a' \neq a, \quad (2.33)$$

where $x \in [0, x_{\max}]$. It follows that this learning rule is risk-averse (as $L_{(a,x)}^v(a)$ is a concave function of x). However, this learning rule is monotonically risk-averse and first-order monotone only if $v_a = v_{a'}$ for all $a, a' \in A$ (because otherwise it fails to be impartial). The analysis of the average return model studied by March [22] which replaces β with $\beta_a = 1/(\kappa_a + 1)$, where κ_a is the number of times action a has been chosen in the past, is similar.

Another learning rule that has received considerable attention is the logistic fictitious play with partial feedback studied by Fudenberg and Levine [17], section 4.8.4. The agent is described by the $|A| \times 2$ matrix (v, κ) where κ_a denotes the number of times action a has been chosen, $\kappa = (\kappa_a)_{a \in A}$, and $v = (v_a)_{a \in A}$ gives the vector of attractions. The next period attraction of an action that was chosen today is its current attraction plus $(x - v_a) / (\kappa_a + 1)$. The attractions of unchosen actions

are not updated. The learning rule is specified as

$$L_{(a,x)}^{v,\kappa}(a) = \frac{e^{v_a+(x-v_a)/(\kappa_a+1)}}{e^{v_a+(x-v_a)/(\kappa_a+1)} + \sum_{a' \neq a} e^{v_{a'}}} \quad (2.34)$$

$$L_{(a',x)}^{v,\kappa}(a) = \frac{e^{v_a}}{e^{v_{a'}+(x-v_{a'})/(\kappa_{a'}+1)} + \sum_{a'' \neq a'} e^{v_{a''}}} \quad \forall a' \neq a \quad (2.35)$$

This learning rule is neither risk-averse nor risk seeking because the curvature of $L_{(a,x)}^{v,\kappa}(a)$ depends on the payoff obtained. Specifically, $L_{(a,x)}^{v,\kappa}(a)$ is concave in x (at σ) for the part of the domain in which $x \geq v_a + (\kappa_a + 1) [\ln(1 - \sigma_a) - \ln \sigma_a]$ and is convex otherwise. Nevertheless, this rule is first-order monotone when $v_a = v_{a'}$ and $\kappa_a = \kappa_{a'}$ for all $a, a' \in A$. The analysis of the risk attitudes of several other learning rules (e.g. minimal information versions of Camerer and Ho [7] and Rustichini [32]), which use a logistic transformation of the attractions to obtain the probabilities with which each action is chosen, is closely related.

F. Discussion

In models of learning probability distributions over actions change from one period to the next in response to some experience. Monotonic risk aversion focuses only on the manner in which the probability of the best actions being chosen changes from one period to the next. This parallels decision theory's focus on the best action. However, for learning, we could look at the entire probability distribution chosen in the next period. Noting that this probability distribution on actions generates a (reduced) distribution over payoffs, which is a weighted average of the payoff distribution of each of the actions, we could ask whether the learning rule is such that expected behavior tomorrow generates a distribution over payoffs which second-order stochastically dominates that of today in every environment. Such a property turns out to be too restrictive. It can be shown that, in environments with

only two actions, the only learning rules which satisfy this condition are the unbiased rules studied by Börgers et al. Unbiased rules exhibit zero expected movement in probability mass when all actions have the same expected payoffs. Such rules satisfy the above condition in a trivial manner because the expected distribution tomorrow is the same as today.⁷ Restricting the set of environments on which the improvement is required would lead us to identify a larger class of learning rules.⁸ We do not pursue this approach in the current analysis.

All the properties we have studied in this chapter have referred to the expected movement of a learning rule. This arises naturally when describing the behavior of the population of individuals each of whom faces the same decision problem. The expected movement of a learning rule has also been studied on many previous occasions when interest has focused on the long run properties of a learning rule. As is well known under conditions of slow learning the actual movement of a learning rule closely approximate it's expected movement.⁹ Combining properties of the expected movement and of the speed of learning inform us about the long term properties of learning rules.

This essay has focused on short term and local properties of learning. Often, however, the long run properties of learning rules are of interest and such an analysis requires us to look at properties that hold globally. That is, the local property would need to hold for each state of learning. This subject, which would require the

⁷It can also be shown that unbiased learning rules are the only learning rules which are continuous in x for all $a, a' \in A$ and satisfy $\sum_a (\sigma_a + f(a)) F_a$ sosd $\sum_a \sigma_a F_a$ in every environment.

⁸For example, we could consider learning rules that satisfy $\sum_a (\sigma_a + f(a)) F_a$ sosd $\sum_a \sigma_a F_a$ in every environment that is completely ordered by the sosd relation. It can be shown that a learning rule is continuous in payoffs and satisfies this condition if and only if it is monotonically risk-averse.

⁹See, for example, Börgers and Sarin [5].

analysis of risk attitudes at each state of learning, is outside the scope of the present study and we leave it for future work. For the examples we discussed in Section E, we are often able to say if the learning rule is globally (monotonically) risk-averse. This is obviously true for the Cross learning rule and the Roth and Erev learning rule which are monotonically risk neutral and monotonically risk-averse at all states, respectively. The weighted return model of March is globally risk-averse though not globally monotonically risk-averse and the logistic fictitious play learning rule is not globally risk-averse.

CHAPTER III

MEAN AND VARIANCE RESPONSIVE LEARNING

A. Introduction

A popular and commonly used measure of the “risk” of a distribution is its variance. Following the seminal papers by Markowitz [23] and Tobin [36] many studies in finance and monetary economics have analyzed alternative distributions of payoffs according to their mean and variance. This literature describes people as seeking higher means and lower variances. If behavior is, to some extent, learnt then it is likely that the learning rules people use lead them to choose actions with higher means and lower variances. In this essay we study such learning rules.¹

We consider learning rules in which the agent has very little prior and feedback information about her environment. She knows the set of actions and chooses each with positive probability.² The agent does not know the distribution of payoffs from any action. In any period, information is obtained about the action chosen and the payoff received. We focus on how this information is used to modify behavior from one period to the next. Since the change in behavior depends on the payoff obtained, the payoff distributions associated with the different actions will determine the expected change in behavior. We refer to this set of distributions as the environment faced by the agent.

Our analysis begins by considering environments in which all actions have the same mean but differ according to their variances. We call a learning rule *variance-*

¹We know of no paper that studies learning rules that respond to both the mean and the variance of a distribution. See Fudenberg and Levine [17] for a comprehensive study of the subject.

²This is a common assumption in learning models. Gul and Pesendorfer [18] provide an axiomatic foundation for such probabilistic choice behavior.

averse if it is expected to result in behavior tomorrow that has a lower variance in every such environment. We call a learning rule *monotonically variance-averse* if the probability with which it chooses the variance minimizing actions is expected to increase in every such environment. We show that variance-averse and monotonically variance-averse rules satisfy a consistency condition that we call *weak unbiasedness* which requires that there is no expected movement in probability on any action if all actions have the same variance (and the same expected payoff). We characterize all weakly unbiased learning rules. All such learning rules allow for a polynomial of order two transformation of payoffs, in which the coefficients of the transformation are allowed to depend on the action chosen and the action whose probability is being updated, before applying Cross' learning rule (Cross [8]).

We characterize all monotonically variance-averse learning rules. Such rules place restrictions on the coefficients of the quadratic transformation of payoffs. We also prove that all monotonically variance-averse learning rules are variance-averse, which in turn we show to be a strict subset of the set of weakly unbiased learning rules.

Our main result considers learning rules in environments in which actions may have different expected payoffs and different variances of payoffs.³ We now study *monotonically mean-variance-averse* rules which require that the probability with which actions which have the highest means and lowest variances, provided they exist, is expected to increase. We characterize monotonically mean-variance-averse rules. Such rules have to be monotonically variance-averse. Additionally, the characterization reveals the restrictions on the relative sizes of the coefficients of the linear and quadratic terms that appear in weakly unbiased learning rules. Intuitively, both

³In such environments, we show that a monotonically variance-averse learning rule may lead to lower expected payoffs.

linear and quadratic terms arise because the learning rule has to respond to both the mean and the variance of the distribution.

We show that monotonically mean-variance-averse rules are expected to move behavior toward lower variance actions in environments in which all actions have the same expected payoff. Furthermore, they move behavior toward higher expected payoffs in environments in which all actions have the same variance. Hence, they capture certain features of rules that respond only to the first moment of payoffs and those that respond only to the second moment of payoffs.

The analysis in this essay complements that in Börgers et al. and the results provided in Chapter II. Börgers et al. characterize *monotonic* learning rules which are expected to increase probability mass on the expected payoff maximizing actions. Chapter II characterizes *monotonically risk-averse* learning rules which are expected to increase probability mass on the actions whose distributions second-order stochastically dominate (sosd) all other actions. In contrast to Börgers et al. which is concerned with the mean of a distribution and the analysis in Chapter II, which is concerned with its risk as measured by sosd, this essay considers learning rules that respond to both the mean of the distribution and its variance, where the latter is used as a measure of risk. Börgers et al. (respectively, the analysis in Chapter II) find that a learning rule is monotonic (respectively, monotonically risk-averse) if and only if the agent transforms the payoff in an affine (respectively, concave) manner. The transformation of payoffs is allowed to depend on the action chosen and the action whose probability is being updated.

Given the results in Börgers et al. and those in Chapter II, the main results provided here may not seem surprising. The arguments involved in it, however, need to be considerably different. This is especially true in the characterization of weakly unbiased learning rules. More importantly, our results allow us to describe agents

whose learning responds to two different features of a distribution. Responding to the first two moments of a distribution seems both to be intuitive and what is observed in experimental data (see, e.g., Erev, Bereby-Meyer and Roth [16] and Haruvy, Erev and Sonsino [20]).

March [22], Denrell and March [13] and Burgos [6] investigate how specific learning rules respond to risk by way of simulations. Denrell [12] analytically considers the long term properties of some learning rules in the same informational setting as this essay. He finds a class of adaptive learning rules that lead to more risk-averse choices in the long run. Also related is Della Vigna and Li Calzi [11] which pays close attention to the long run risk attitudes of a specific learning rule. The learning rule they consider, however, assumes more knowledge about the environment than the rules we consider. Finally, the work of Karni and Schmeidler [21] and Robson [27], [28] is related. In their work current preferences over gambles, i.e., risk attitudes, are the result of an evolutionary process in which either the expected number of offspring or the probability of survival is maximized.

This essay is structured as follows. In the next section we introduce the framework. Section C and Section D consider the case in which all actions have the same expected payoff. Section C characterizes learning rules which satisfy a weak consistency requirement that all variance-averse and monotonically variance-averse learning rules need to satisfy. Section D characterizes monotonically variance-averse rules and provides some necessary and some sufficient conditions for variance-averse rules. In Section E we consider environments in which actions have different means and variances of payoff and provide results concerning monotonically mean-variance-averse rules. Section F provides some concluding comments.

B. Framework

The decision maker has a finite number of actions $a \in A$. Payoffs are interpreted as monetary magnitudes that lie in some bounded interval known to the decision maker which we normalize to be $[0, 1]$.⁴ For any subset $D \subset [0, 1]$ the probability that a gives a payoff in D is $\mu_a(D)$.⁵ Let $\mu = (\mu_a)_{a \in A}$ denote the vector of probability measures associated with each action. We refer to $E \equiv (A, \mu)$ as an environment. The expected payoff from action a is denoted $\pi_a = \int_0^1 x d\mu_a$ and $A^* := \{a \in A : \pi_a \geq \pi_{a'} \forall a' \in A\}$ denotes the set of expected payoff maximizing actions. Let $\xi_a = \int_0^1 x^2 d\mu_a$ denote the second moment of payoffs of action a . The variance of payoffs of action a is given by $\rho_a = \xi_a - \pi_a^2$. Let $A_* := \{a \in A : \rho_a \leq \rho_{a'} \forall a' \in A\}$ denote the set of variance minimizing actions.

The individual knows the set of actions A but does not know μ . She chooses among her actions today according to the mixed action vector $\sigma \in \Delta(A)$, where $\Delta(A)$ is the set of all probability distributions over A . σ describes the behavior of the individual today. We assume that σ lies in the interior of the simplex $\Delta(A)$.⁶ The expected payoff associated with σ is $\pi_\sigma = \sum_a \sigma_a \pi_a$ and the variance associated with it is $\rho_\sigma = \sum_a \sigma_a \xi_a - \pi_\sigma^2$.

A learning rule specifies the individuals behavior tomorrow as a function of her behavior today, the action she chooses and the payoff she obtains. The learning rule should be viewed as a “reduced form” of the true learning process. The true learning rule may, for example, specify how the decision maker updates her beliefs and how these beliefs are transformed into behavior. Combining the two steps of

⁴This normalization is without loss of generality for our results.

⁵The probability that action a gives a payoff x is denoted by $\mu_a(x)$.

⁶It is easy to show that the properties of learning rules investigated in this essay cannot be satisfied if this condition is violated.

behavior updating and belief change we get the learning rule we specify. We shall assume the behavior today is given and fixed and hence we may write a learning rule as $L : A \times [0, 1] \rightarrow \Delta(A)$.

Denote by $L_{(a',x)}(a)$ the probability that the learning rule L assigns to action a tomorrow if action a' was chosen today and a payoff of x is received. Fix an environment and a learning rule. Then, the expected change in the probability of action a is given by $f(a) = \sum_{a' \in A} \sigma_{a'} \int_0^1 L_{(a',x)}(a) d\mu_{a'} - \sigma_a$. For any subset of actions, $\tilde{A} \subset A$, let $f(\tilde{A})$ denote the expected change in probability mass on that subset. That is, $f(\tilde{A}) = \sum_{a \in \tilde{A}} f(a)$. The expected change in expected payoffs is given by $g = \sum_a f(a) \pi_a$. The expected change in the variance of the payoff is given by $h = \sum_{a \in A} f(a) \xi_a - [\mathbf{E}(\pi_{\sigma'})^2 - (\pi_{\sigma})^2]$, where σ' is the next period behavior and the expectation \mathbf{E} is taken with regard to its value.

We shall begin our analysis by studying environments in which each action gives the same expected payoff which we refer to as flat environments.

Definition 8 *An environment is flat if $\pi_a = \pi_{a'}$ for all $a, a' \in A$.*

In flat environments $A = A^*$ and $h = \sum_a f(a) \xi_a$. We begin our analysis with flat environments because this allows us to isolate the effect of (payoff) variance on the learning rule. We now introduce two properties of learning rules which pertain to their behavior in flat environments.

Definition 9 *A learning rule is variance-averse (/neutral/seeking/) if for all flat environments with $A \neq A_*$, $h < (/ = / > /) 0$.*

In words, a learning rule is said to be variance-averse (/neutral/seeking/) if in all environments in which each action has the same expected payoff, but at least two actions have a different variance of payoff, the expected change in the variance

of payoff is strictly negative (/zero/strictly positive/). Hence, a learning rule is variance-averse if it is expected to lead to a strict reduction in the variance of payoff from one period to the next, provided the environment stays the same.

Definition 10 *A learning rule is monotonically variance-averse (/neutral/seeking/) if in all flat environments with $A \neq A_*$, $f(A_*) > (/ = / < /) 0$.*

In words, a learning rule is monotonically variance-averse (/neutral/seeking/) if in all environments in which each action has the same expected payoff, but at least two actions have a different variance of payoff, the expected change in the probability with which the variance minimizing actions are chosen is strictly positive (/zero/strictly negative/).⁷

Remark 1 *If $|A| = 2$ then a learning rule is variance-averse (/neutral/seeking/) if and only if the leaning rule is monotonically variance-averse (/neutral/seeking/).*

We end this section with an example of a learning rule due to Cross (1973).

Example 1 *Cross Learning rule*

$$L_{(a,x)}(a) = \sigma_a + (1 - \sigma_a)x \quad (3.1)$$

$$L_{(a,x)}(a') = \sigma_{a'} - \sigma_{a'}x \quad \forall a' \neq a \quad (3.2)$$

In the Cross learning rule,

$$f(a) = \sigma_a(\pi_a - \sum_{a'} \sigma_{a'} \pi_{a'}) \quad (3.3)$$

⁷An alternate, and perhaps more intuitive way to define monotonically variance seeking rules, could be as those learning rules that satisfy $f(A_{*-}) > 0$, where $A_{*-} := \{a \in A : \rho_a \geq \rho_{a'} \forall a' \in A\}$ in every environment with $A \neq A_*$. It turns out that the learning rules for which $f(A_{*-}) > 0$ are precisely those for which $f(A_*) < 0$ for all flat environments with $A \neq A_*$.

for all $a \in A$. In flat sets, $f(a) = 0$ for all a and

$$h = \sum_{a \in A} \sigma_a \xi_a (\pi_a - \sum_{a'} \sigma_{a'} \pi_{a'}) = 0. \quad (3.4)$$

Hence, this rule is variance neutral and monotonically variance neutral.

C. Weak Unbiasedness

We begin this section by defining the concept of unbiasedness which was introduced in Börger et al. Unbiasedness requires that in all flat environments there is no expected movement in probability mass.

Definition 11 *A learning rule L is said to be unbiased if in all flat environment $f(a) = 0$ for each $a \in A$.*

The following remark, which is a straightforward consequence of the definitions, reveals a connection between unbiased learning rules and those that are variance neutral or monotonically variance neutral. It also generalizes the observation in the preceding section that the Cross learning rule, which is unbiased, is variance neutral and monotonically variance neutral.

Remark 2 *Every unbiased learning rule is variance neutral and monotonically variance neutral.*

Next we introduce a property related to unbiasedness which we call weak unbiasedness. This requires that in environments in which all actions have the same expected payoff and the same variance of payoff there is no expected movement in probability mass. Lemma 4 below shows why we are interested in weakly unbiased learning rules.

Definition 12 *A learning rule L is weakly unbiased if in all flat environments with $A = A_*$, $f(a) = 0$ for all $a \in A$.*

Lemma 4 *Every learning rule that is variance-averse (/neutral/seeking/) or monotonically variance-averse (/neutral/seeking/) is weakly unbiased.*

Proof. We first prove this result for variance-averse case in Step 1. The proof is by contradiction. Suppose the learning rule L is variance-averse and monotonically variance-averse but that it is not weakly unbiased. Hence, there exist a flat environment E with $A = A_*$ and $f(a) > 0$ for some $a \in A$.

Step 1: We consider two cases:

Case 1: $\mu_a((0, 1)) > 0$. Consider another environment \tilde{E} in which action a removes ε probability uniformly from $(0, 1)$ ⁸ and assigns $(1 - \pi)\varepsilon$ to 0 and $\pi\varepsilon$ to 1. Other than this difference both environments offer the same distributions of payoffs for all actions. By construction, the payoff distribution of a in \tilde{E} is a mean preserving spread of a in E . Let $\tilde{f}(a)$ denote the expected change in probability on a in environment \tilde{E} . Since $\tilde{f}(a)$ is continuous in ε , there is a small enough ε such that $\tilde{f}(a) > 0$. This contradicts the assumption that L is variance-averse and monotonically variance-averse.

Case 2: $\mu_a(0, 1) = 0$. We break down the argument into three subcases.

Subcase 2.1: $\mu_a(1) = 1$. Since the environment is flat all actions $a' \in A$ must also have $\mu_{a'}(1) = 1$. Consider another environment \hat{E} in which each action $a \in A$ assigns probability $(1 - \varepsilon)$ to 1 and ε to $y \in (0, 1)$. Now consider an environment \tilde{E} in which all actions except a have the same payoff distributions as in \hat{E} and a removes probability ε from y and assigns $\varepsilon/2$ to $y - \delta$ and $\varepsilon/2$ to $y + \delta$ where $0 < \delta < \min\{1 - y, y\}$. By construction, a in \tilde{E} is a mean preserving spread of

⁸By which we mean that for every subset $D \subset (0, 1)$, $\tilde{\mu}(D) = (1 - \varepsilon)\mu(D)$.

a in \hat{E} . As $\tilde{f}(a)$ is continuous in ε there is a small enough ε such that $\tilde{f}(a) > 0$. This contradicts the assumption that L is variance-averse and monotonically variance-averse.

Subcase 2.2: $\mu_a(0) = 1$. The argument is analogous to Subcase 2.1.

Subcase 2.3: $\mu_a(1) \neq 1, \mu_a(0) \neq 1, \mu_a(0) + \mu_a(1) = 1$. In this case, $\pi_a \in (0, 1)$. Now consider an environment \hat{E} in which each $a \in A$ has $\hat{\mu}_a(1) = (1 - \varepsilon)\mu_a(1)$ and $\hat{\mu}_a(0) = (1 - \varepsilon)\mu_a(0)$ and $\hat{\mu}_a(\pi_a) = \varepsilon$. Now consider an environment \tilde{E} in which all actions except a have the same payoff distributions as in \hat{E} and a removes probability ε from π_a and assigns $\varepsilon/2$ to $\pi_a + \delta$ and $\varepsilon/2$ to $\pi_a - \delta$ for some $\delta \in (0, \min\{\pi, (1 - \pi)\})$. By construction, the distribution of a in \tilde{E} is a mean preserving spread of a in \hat{E} . Since $f(a) > 0$ by hypothesis, and because $\tilde{f}(a)$ is continuous in ε we have that for small ε , $\tilde{f}(a) > 0$. This contradicts the assumption that L is variance-averse and monotonically variance-averse.

Step 2: The above argument suffices to show that every variance neutral and monotonically variance neutral rule is weakly unbiased. One simply has to replace “variance-averse” with “variance neutral” in the above argument.

Step 3: The proof when a learning rule is variance seeking and monotonically variance seeking is analogous to that provided in Step 1. In this case we begin by considering a flat environment E with $A = A_*$ and supposing that $f(a) < 0$. We can then construct a new environment \tilde{E} in which all actions $a' \neq a$ have the same distribution as in E but action a in \tilde{E} has a distribution which is a mean preserving spread of the distribution of a in E . Continuity of f function will ensure that for a small enough change in the the distribution of a in E , we will have $\tilde{f}(a) < 0$ contradicting that the rule is variance (and monotonically variance) seeking. \square

The next result characterizes the set of weakly unbiased learning rules.

Proposition 4 *A learning rule is weakly unbiased if and only if there exist matrices $(A_{aa'})_{a,a'=1,\dots,|A|}$, $(B_{aa'})_{a,a'=1,\dots,|A|}$ and $(C_{aa'})_{a,a'=1,\dots,|A|}$ such that for every $(a, x) \in A \times [0, 1]$*

$$L_{(a,x)}(a) = \sigma_a + (1 - \sigma_a) (A_{aa} + B_{aa}x - C_{aa}x^2) \quad (3.5)$$

$$L_{(a,x)}(a') = \sigma_{a'} - \sigma_{a'} (A_{aa'} + B_{aa'}x - C_{aa'}x^2) \quad \forall a' \neq a \quad (3.6)$$

and for all a ,

$$A_{aa} = \sum_{a'} \sigma_{a'} A_{a'a} \quad (3.7)$$

$$B_{aa} = \sum_{a'} \sigma_{a'} B_{a'a} \quad (3.8)$$

$$C_{aa} = \sum_{a'} \sigma_{a'} C_{a'a} \quad (3.9)$$

Proof.

Sufficiency

Consider a flat environment E with $A = A_*$ and suppose the learning rule satisfying the properties stated in Proposition 4. Then (3.5) – (3.6) imply that, for all a ,

$$f(a) = \sigma_a \int_0^1 (1 - \sigma_a) (A_{aa} + B_{aa}x - C_{aa}x^2) d\mu_a \quad (3.10)$$

$$- \sum_{a' \neq a} \sigma_{a'} \int_0^1 \sigma_a (A_{a'a} + B_{a'a}x - C_{a'a}x^2) d\mu_{a'} \quad (3.11)$$

$$= \sigma_a \left[A_{aa} + B_{aa}\pi_a - C_{aa}\xi_a - \sum_{a'} \sigma_{a'} (A_{a'a} + B_{a'a}\pi_{a'} - C_{a'a}\xi_{a'}) \right]. \quad (3.12)$$

Equations (3.7) – (3.9) and the fact that $A = A^* = A_*$ imply $f(a) = 0$ for all a .

Necessity:

Let L be a weakly unbiased learning rule and let E and \tilde{E} be two flat environments with $A = A_*$. Furthermore, suppose the variance of payoff ρ is strictly between

the minimum ρ_{\min} and maximum ρ_{\max} possible. The distribution for all actions a in E is given by $(0, x, \pi_a, 1; \mu_a(0), \mu_a(x), \mu_a(\pi_a), \mu_a(a))$.⁹ In environment \tilde{E} all actions $a' \neq a$ have the same distribution as in E , while action a has a distribution given by $(0, \pi, 1; \tilde{\mu}_a(0), \tilde{\mu}_a(\pi), \tilde{\mu}_a(1))$. By assumption, $\tilde{\pi}_a = \pi_a$ and $\tilde{\rho}_a = \rho_a$.

To see that for any mean π , the variance ρ (with $\rho_{\min} < \rho < \rho_{\max}$) can be achieved by a suitably chosen distribution with support on the payoffs 0, π and 1, we need to observe that the probabilities $\tilde{\mu}_a(0)$, $\tilde{\mu}_a(\pi)$, and $\tilde{\mu}_a(1)$ have to satisfy the three equations $\tilde{\mu}_a(0) + \tilde{\mu}_a(\pi) + \tilde{\mu}_a(1) = 1$, $\tilde{\mu}_a(0) \cdot 0 + \tilde{\mu}_a(\pi) \cdot \pi + \tilde{\mu}_a(1) \cdot 1 = \pi$ and $\tilde{\mu}_a(0) \cdot 0^2 + \tilde{\mu}_a(\pi) \cdot \pi^2 + \tilde{\mu}_a(1) \cdot 1^2 = \xi$. These three equations can be solved to give,

$$\tilde{\mu}_a(0) = \frac{\xi - \pi^2}{\pi} \quad (3.13)$$

$$\tilde{\mu}_a(\pi) = \frac{\pi - \xi}{\pi - \pi^2} \quad (3.14)$$

$$\tilde{\mu}_a(1) = \frac{\xi - \pi^2}{1 - \pi}, \quad (3.15)$$

which have a unique interior solution between 0 and 1 given $\rho_{\min} < \rho < \rho_{\max}$.

Next we solve the three equations $\mu_a(0) + \mu_a(x) + \mu_a(\pi) + \mu_a(1) = 1$, $\mu_a(0) \cdot 0 + \mu_a(x) \cdot x + \mu_a(\pi) \cdot \pi + \mu_a(1) \cdot 1 = \pi$ and $\mu_a(0) \cdot 0^2 + \mu_a(x) \cdot x^2 + \mu_a(\pi) \cdot \pi^2 + \mu_a(1) \cdot 1^2 = \xi$ in terms of $\mu_a(x)$ to obtain

$$\mu_a(0) = \frac{\xi - \pi^2}{\pi} - \mu_a(x) \left(1 - \frac{x - x^2}{\pi - \pi^2} - \frac{\pi x^2 - \pi^2 x}{\pi - \pi^2} \right) \quad (3.16)$$

$$\mu_a(\pi) = \frac{\pi - \xi}{\pi - \pi^2} - \mu_a(x) \frac{x - x^2}{\pi - \pi^2} \quad (3.17)$$

$$\mu_a(1) = \frac{\xi - \pi^2}{1 - \pi} - \mu_a(x) \frac{\pi x^2 - \pi^2 x}{\pi - \pi^2}. \quad (3.18)$$

⁹The first four terms in the brackets are payoffs and the last four terms are their respective probabilities.

Substituting (3.13) in (3.16), (3.14) in (3.17) and (3.15) in (3.18) we get,

$$\mu_a(0) = \tilde{\mu}_a(0) - \mu_a(x) \left(1 - \frac{x - x^2}{\pi - \pi^2} - \frac{\pi x^2 - \pi^2 x}{\pi - \pi^2} \right) \quad (3.19)$$

$$\mu_a(\pi) = \tilde{\mu}_a(\pi) - \mu_a(x) \frac{x - x^2}{\pi - \pi^2} \quad (3.20)$$

$$\mu_a(1) = \tilde{\mu}_a(1) - \mu_a(x) \frac{\pi x^2 - \pi^2 x}{\pi - \pi^2}. \quad (3.21)$$

Let $\bar{L}_{(a,.)}(a')$ denote the expected value of $\sigma_{a'}$ tomorrow given that a was chosen today. That is, $\bar{L}_{(a,.)}(a') = \int_0^1 L_{(a,x)}(a') d\mu_a$. We now compute the expected change in the probability of an arbitrary action a' in the two environments,

$$\begin{aligned} f(a') &= \sigma_a [\mu_a(0) L_{(a,0)}(a') + \mu_a(x) L_{(a,x)}(a') + \mu_a(\pi) L_{(a,\pi)}(a') \\ &\quad + \mu_a(1) L_{(a,1)}(a') + \sum_{a'' \neq a} \sigma_{a''} \bar{L}_{(a'',.)}(a') - \sigma_{a'} = 0 \end{aligned} \quad (3.22)$$

$$\begin{aligned} \tilde{f}(a') &= \sigma_a [\tilde{\mu}_a(0) L_{(a,0)}(a') + \tilde{\mu}_a(\pi) L_{(a,\pi)}(a') + \tilde{\mu}_a(1) L_{(a,1)}(a')] \\ &\quad + \sum_{a'' \neq a} \sigma_{a''} \bar{L}_{(a'',.)}(a') - \sigma_{a'} = 0. \end{aligned} \quad (3.23)$$

Subtracting these two equations and slightly rearranging we get,

$$\begin{aligned} \mu_a(x) L_{(a,x)}(a') &= [\tilde{\mu}_a(0) - \mu_a(0)] L_{(a,0)}(a') \\ &\quad + [\tilde{\mu}_a(\pi) - \mu_a(\pi)] L_{(a,\pi)}(a') \\ &\quad + [\tilde{\mu}_a(1) - \mu_a(1)] L_{(a,1)}(a') \end{aligned} \quad (3.24)$$

Substituting (3.19) – (3.21) in (3.24) and slightly re-arranging we obtain,

$$\begin{aligned} L_{(a,x)}(a') &= \left[1 - \frac{x - x^2}{\pi - \pi^2} - \frac{\pi x^2 - \pi^2 x}{\pi - \pi^2} \right] L_{(a,0)}(a') \\ &\quad + \frac{x - x^2}{\pi - \pi^2} L_{(a,\pi)}(a') \\ &\quad + \frac{\pi x^2 - \pi^2 x}{\pi - \pi^2} L_{(a,1)}(a'), \end{aligned} \quad (3.25)$$

which is a polynomial of order no greater than two in x . Hence, $L_{(a,x)}(a')$ can be written as

$$L_{(a,x)}(a') = \hat{A}_{aa'} + \hat{B}_{aa'}x + \hat{C}_{aa'}x^2. \quad (3.26)$$

Straightforward algebra reveals that these conditions impose no extra restrictions on $\hat{A}_{aa'}, \hat{B}_{aa'}, \hat{C}_{aa'}$ beyond $\hat{A}_{aa'} = L_{(a,0)}(a')$. In particular, these coefficients do not depend on π .

Equating,

$$L_{(a,x)}(a) = \hat{A}_{aa} + \hat{B}_{aa}x + \hat{C}_{aa}x^2 \quad (3.27)$$

with (3.5), we set the coefficients of A_{aa}, B_{aa} and C_{aa} in such a way that they satisfy

$$\begin{aligned} \hat{A}_{aa} &= \sigma_a + (1 - \sigma_a) A_{aa} \\ \hat{B}_{aa} &= (1 - \sigma_a) B_{aa} \\ \hat{C}_{aa} &= -(1 - \sigma_a) C_{aa}. \end{aligned}$$

Similarly, we set the coefficients of $A_{aa'}, B_{aa'}$ and $C_{aa'}$ in such a way that they satisfy

$$\begin{aligned} \hat{A}_{a'a} &= \sigma_a - \sigma_a A_{a'a} \\ \hat{B}_{a'a} &= -\sigma_a B_{a'a} \\ \hat{C}_{a'a} &= \sigma_a C_{a'a}. \end{aligned}$$

Finally, to prove the necessity of conditions (3.7)–(3.9) consider an environment in which every action pays x with probability 1. In this environment

$$f(a) = \sigma_a \left[(A_{aa} + B_{aa}x + C_{aa}x^2) - \sum_{a'} \sigma_{a'} (A_{a'a} + B_{a'a}x + C_{a'a}x^2) \right]. \quad (3.28)$$

This expression has to be zero for all $x \in [0, 1]$. This requires that conditions (3.7)–(3.9) are satisfied. \square

Conditions (3.5) and (3.6) of Proposition 1 reveal that all weakly unbiased rules apply Cross' rule after transforming payoffs by a polynomial of order two in which the coefficients of the transformation are allowed to depend on the chosen action and the action whose probability is being updated. Conditions (3.7)-(3.9) restrict the coefficients of this transformation.

Remark 3 *The set of unbiased learning rules identified by Börgers et al. is contained in the set of weakly unbiased learning rules. They are obtained by setting $C_{a'a} = 0$ for all $a, a' \in A$.*

Remark 4 *The expected change in probability for an action a for all weakly unbiased learning rules is given by*

$$f(a) = \sigma_a \sum_{a'} \sigma_{a'} [B_{a'a}(\pi_a - \pi_{a'}) - C_{a'a}(\xi_a - \xi_{a'})]. \quad (3.29)$$

which in a flat environment reduces to

$$f(a) = \sigma_a \sum_{a'} \sigma_{a'} C_{a'a}(\xi_{a'} - \xi_a). \quad (3.30)$$

As Remark 4 shows, the coefficients $C_{a'a}$ play a fundamental role in determining the expected movement of probability in flat sets. The precise relation between these coefficients and attitudes towards variance of learning rules is studied in the next section.

D. Learning and Behavior Toward Variance

The next Lemma provides restrictions on the C_{aa} coefficients for a learning rule to be variance-averse (/neutral/seeking/) or monotonically variance-averse (/neutral/seeking/).

Lemma 5 *Every variance-averse (/neutral/seeking/) learning rule and every monotonically variance-averse (/neutral/seeking/) learning rule has $C_{aa} > (/ = / < /) 0$.*

Proof. Consider a flat environment in which $\mu_a(x) = 1$ for some $x \in (0, 1)$, and all other actions $a' \neq a$ have $\mu_{a'}(x) = 1 - \varepsilon$, and $\mu_{a'}(x - \delta) = \varepsilon/2$ and $\mu_{a'}(x + \delta) = \varepsilon/2$ (for $0 < \delta < \min\{x, 1 - x\}$). Clearly, $\xi_a = x^2$ and $\xi_{a'} = x^2 + \varepsilon\delta^2$ and $A_* = \{a\}$. Furthermore,

$$f(a) = \sigma_a \left[\sum_{a' \neq a} \sigma_{a'} C_{a'a} (\xi_{a'} - \xi_a) \right] \quad (3.31)$$

$$= \sigma_a \left[\sum_{a' \neq a} \sigma_{a'} C_{a'a} \varepsilon \delta^2 \right] \quad (3.32)$$

$$= \sigma_a (1 - \sigma_a) C_{aa} \varepsilon \delta^2. \quad (3.33)$$

Hence, for a learning rule to be variance-averse (/neutral/seeking) or monotonically variance-averse (/neutral/seeking) we require $C_{aa} > (/ = / < /) 0$. \square

From Lemma 2 it follows that no variance-averse or monotonically variance-averse learning rule is Börgers et al. monotone because such learning rules cannot be Börgers et al. unbiased (see Remark 3).

Proposition 5 *A learning rule is monotonically variance-averse (/neutral/seeking/) if and only if*

1. $C_{a'a} \geq (/ = / \leq /) 0$ for all $a' \neq a$,
2. if D is a non-empty proper subset of A then there are actions $a \in D$ and $a' \in A - D$ such that $C_{a'a} > (/ = / < /) 0$.

Proof. *Sufficiency:*

From Lemma 4 and Proposition 4 we know that every monotonically variance-averse (/neutral/seeking/) learning rule has to satisfy equations (3.5) – (3.9). (3.5)

and (3.6) imply

$$f(a) = \sigma_a \int_0^1 (1 - \sigma_a) (A_{aa} + B_{aa}x - C_{aa}x^2) d\mu_a \quad (3.34)$$

$$- \sum_{a' \neq a} \sigma_{a'} \int_0^1 \sigma_a (A_{a'a} + B_{a'a}x - C_{a'a}x^2) d\mu_{a'} \quad (3.35)$$

$$= \sigma_a \left[(A_{aa} + B_{aa}\pi_a - C_{aa}\xi_a) - \sum_{a'} \sigma_{a'} (A_{a'a} + B_{a'a}\pi_{a'} - C_{a'a}\xi_{a'}) \right] \quad (3.36)$$

Since the environment is flat, and using (3.7) – (3.9), we get

$$f(a) = \sigma_a \left[\sum_{a'} \sigma_{a'} C_{a'a} (\xi_{a'} - \xi_a) \right]. \quad (3.37)$$

Since $\xi_{a'} - \xi_a > 0$ for every pair of actions $a \in A_*$ and $a' \in A - A_*$, $f(A_*) > (/ = / < /) 0$ by the restrictions on C stated in the Proposition.

Necessity:

We focus on the case of variance aversion (the arguments for the other cases are analogous). The proof is by contradiction. Suppose condition 1 is not satisfied so that $C_{a'a} < 0$ for some $a' \neq a$. Consider a flat environment in which a gives a payoff of $x \in (0, 1)$ with probability one, and a' gives x with probability $1 - \varepsilon$ and $x + \delta$ with probability $\varepsilon/2$ and gives $x - \delta$ with probability $\varepsilon/2$, where $\delta \in (0, \min\{x, 1 - x\})$ and $\varepsilon \in (0, 1]$. All other actions a'' , if any, give a payoff of $x \in (0, 1)$ with probability $1 - \varepsilon'$ and $x + \delta$ with probability $\varepsilon'/2$ and gives $x - \delta$ with probability $\varepsilon'/2$ where $\varepsilon' \in (0, 1]$. It is easily seen that $\xi_a = x^2$, $\xi_{a'} = x^2 + \varepsilon\delta^2$, $\xi_{a''} = x^2 + \varepsilon'\delta^2$ and $A_* = \{a\}$. Hence, the conditions on L derived in Proposition 4 imply that

$$f(a) = \sigma_a \left\{ \sigma_a C_{a'a} (\xi_{a'} - \xi_a) + \sum_{a'' \neq a, a'} \sigma_{a''} C_{a''a} (\xi_{a''} - \xi_a) \right\} \quad (3.38)$$

$$= \sigma_a \left\{ \sigma_a C_{a'a} \varepsilon \delta^2 + \sum_{a'' \neq a, a'} \sigma_{a''} C_{a''a} \varepsilon' \delta^2 \right\}. \quad (3.39)$$

For ε' small enough $f(a) = f(A_*) < 0$ which contradicts monotonicity. Therefore, $C_{a'a}$ cannot be less than zero for any a, a' with $a \neq a'$

Next, suppose condition 2 is not satisfied. Then there exists a non-empty subset of $D \subset A$, such that $A - D \neq \emptyset$ and $C_{a'a} = 0$ for all pairs of lotteries $a \in D$ and $a' \in A - D$. Now consider an environment in which $a \in D$ receives a payoff of x with probability 1 and each action $a' \in A - D$ receives x with probability $1 - \varepsilon$ and receives $x - \delta$ with probability $\varepsilon/2$ and receives $x + \delta$ with probability $\varepsilon/2$ where $\varepsilon \in (0, 1]$ and $\delta \in (0, \min\{x, 1 - x\})$. Clearly, $A_* = D$. Also, for any $a \in D$, the restrictions of Proposition 1 imply

$$f(a) = \sigma_a \sum_{a' \in A - D} \sigma_{a'} C_{a'a} (\xi_{a'} - \xi_a). \quad (3.40)$$

The RHS vanishes because of our assumption that $C_{a'a} = 0$ for all actions $a' \in A - D$ and $a \in D$. Therefore, $f(a) = f(A_*) = 0$ which contradicts the hypothesis that the rule is monotonically variance-averse. \square

Our next result shows that every learning rule that is monotonically variance-averse (/neutral/seeking/) is variance-averse (/neutral/seeking/).

Proposition 6 *Every monotonically variance-averse (/neutral/seeking/) learning rule is variance-averse (/neutral/seeking/).*

Proof. The proof is analogous to the one provided in Börgers et al. for their Proposition 3(ii). In our case, we have to perform the induction on the second moments rather than on the first moments. We omit the details. \square

It is easy to see that monotone variance-averse rules are a very special case of the monotonically risk-averse rules introduced in Chapter II. This specific class is the one in which the response towards the risk of a distribution is completely determined by

the response toward its variance.¹⁰ The relationship between monotonically variance-averse rules and the monotonically risk-averse learning rules studied in Chapter II is analogous to the relation between quadratic utility functions and risk-averse utility functions in expected utility theory (Meyer [24]).

The next example motivates us to consider learning rules that respond in desirable ways to both higher mean and lower variance.

Example 2 *Negative Cross-Square Learning Rule*

$$L_{(a,x)}(a) = \sigma_a + (1 - \sigma_a)(1 - x^2) \quad (3.41)$$

$$L_{(a,x)}(a') = \sigma_{a'} - \sigma_{a'}(1 - x^2) \quad \forall a' \neq a. \quad (3.42)$$

The negative Cross-Square rule has $A_{a'a} = C_{a'a} = 1$ and $B_{a'a} = 0$ for all $a, a' \in A$. Hence, this learning rule is variance-averse and monotonically variance-averse. Note, however, that the higher the payoff obtained the more probability mass that is taken away from the chosen action. This has the consequence that this rule may not have desirable properties on non-flat sets. In particular, it is not difficult to construct an example in which each action gives a unique payoff (with probability one) and the action with the highest payoff has a negative expected movement.

E. Monotonically Mean-Variance-Averse Rules

In environments that are not necessarily flat, we now define the most desirable set of actions for an agent who prefers higher means and lower variances. Specifically, we now consider $A_*^* = \{a : \pi_a \geq \pi_{a'} \wedge \rho_a \leq \rho_{a'} \text{ for all } a' \in A\}$ which defines the

¹⁰From Remark 4 it follows that in flat environments $f(a) = \sigma_a \sum_{a'} \sigma_{a'} C_{a'a} (\rho_{a'} - \rho_a)$. So, only the variance matters in determining expected movement.

set of actions that have the maximum expected payoff and the minimum variance of payoff.

Definition 13 *A learning rule is monotonically mean-variance-averse if $f(A_*) > 0$ in all environments with $A_*^* \neq A$ and $A_*^* \neq \emptyset$.*

It is easy to see that every monotonically mean-variance-averse learning rule is monotonically variance-averse. Hence, every monotonically mean-variance-averse learning rule satisfies the conditions we derived on monotonically variance-averse rules. The next Proposition provides the necessary and sufficient conditions for a learning rule to be monotonically mean-variance-averse. As will be seen this provides some additional restrictions on the relative magnitudes of the B and C coefficients.

Proposition 7 *A learning rule is monotonically mean-variance-averse if and only if*

1. *The learning rule is monotonically variance-averse.*
2. *$B_{a'a} - 2C_{a'a}x \geq 0$ for all $x \in (0, 1)$ and for all $a, a' \in A$.*
3. *If D is a nonempty subset of A , then for all $x \in (0, 1)$ there exists an action $a \in D$ and an action $a' \in A - D$ such that $B_{a'a} - 2C_{a'a}x > 0$.*

Proof. *Sufficiency:*

Let $a \in A_*^*$. Then,

$$f(a) = \sigma_a \left\{ (B_{aa}\pi_a - C_{aa}\xi_a) - \sum_{a'} \sigma_{a'} (B_{a'a}\pi_{a'} - C_{a'a}\xi_{a'}) \right\}. \quad (3.43)$$

Conditions 2 and 3 of the hypothesis and conditions 1 and 2 in Proposition 5 ensure that at least one of the terms is strictly positive for some $a \in A_*^*$.

Necessity:

The necessity of condition 1 in this Proposition is obvious. To see that condition 2 is necessary suppose by way of contradiction that $B_{a'a} - 2C_{a'a}x < 0$ for some $a' \neq a$

and some $x \in (0, 1)$. Consider an environment in which $\mu_a(x) = 1$ so that $\pi_a = x$ and $\rho_a = 0$ for some $x \in (0, 1)$. Furthermore, suppose that $\mu_{a'}(x - \delta) = 1$ for $\delta \in (0, x)$ so that $\pi_{a'} = x - \delta$ and $\rho_{a'} = 0$ for some $x \in (0, 1)$. Lastly, suppose that for any a'' we have $\mu_{a''}(x - \varepsilon) = 1$ for $\varepsilon \in (0, x)$ so that $\pi_{a''} = x - \varepsilon$ and $\rho_{a''} = 0$ for some $x \in (0, 1)$. Clearly, in this environment $A_*^* = \{a\} \neq A$. We have

$$f(a) = \sigma_a \left\{ \sum_{a'} \sigma_{a'} [(B_{a'a}\pi_a - C_{a'a}\pi_a^2) - (B_{a'a}\pi_{a'} - C_{a'a}\pi_{a'}^2)] \right\} \quad (3.44)$$

$$= \sigma_a \{ \sigma_{a'} [(B_{a'a}x - C_{a'a}x^2) - (B_{a'a}(x - \delta) - C_{a'a}(x - \delta)^2)] \} \quad (3.45)$$

$$+ \sum_{a''} \sigma_{a''} [(B_{a''a}x - C_{a''a}x^2) - (B_{a''a}(x - \varepsilon) - C_{a''a}(x - \varepsilon)^2)] \} \quad (3.46)$$

For small enough δ the first term in the first square bracket is negative. Hence, for small enough ε , we have $f(a) < 0$ which contradicts the learning rule being monotonically mean-variance-averse.

Finally, to prove the necessity of condition 3, suppose by way of contradiction that for some $x \in (0, 1)$ we have $B_{a'a} - 2C_{a'a}x = 0$ for all a, a' with $a \in D$ and $a' \in A - D$. We know from the necessity of condition 2 of Proposition 2 that $C_{a'a} > 0$ for some $a \in D$ and $a' \in A - D$. This implies that there exists an $x' \in (x, 1)$ such that $B_{a'a} - 2C_{a'a}x' < 0$ which violates condition 2 and therefore violates mean-variance aversion. \square

Monotone mean-variance-averse rules have several useful properties. The next result reveals that such rules are expected to result in behavior that reduces variance in flat sets. In environments in which all actions have the same variance such rules tend to increase expected payoffs.

Proposition 8 *Suppose L is a monotonically mean-variance-averse rule. Then:*

1. *In every flat set with $A_*^* \neq A$, we have $h < 0$.*
2. *In every set with $\rho_a = \rho_{a'}$ for all $a, a' \in A$ and $A_*^* \neq A$, we have $g > 0$.*

Proof. The argument follows from an inductive argument similar to that provided in Proposition 3(ii) in Börgers et al. We omit the details. \square

We end this section with an example of a learning rule that is monotonically mean-variance-averse.

Example 3 *Mean-Variance Cross Learning rule*

$$L_{(a,x)}(a) = \sigma_a + (1 - \sigma_a) \left(x - \frac{1}{2}x^2 \right) \quad (3.47)$$

$$L_{(a,x)}(a') = \sigma_{a'} - \sigma_a \left(x - \frac{1}{2}x^2 \right) \quad \forall a' \neq a. \quad (3.48)$$

Comparing Example 2 with Example 3 we see that in the Mean-Variance Cross learning rule the updated probability of the chosen action responds positively to the obtained payoff whereas in the Negative Cross learning rule this response is negative. Intuitively, this difference allows the Mean-Variance Cross learning rule to respond positively to the expected payoff of the chosen action.

F. Conclusion

In the class of learning rules in which people learn from their experience, we have characterized those that would be expected to lead them to choose actions with higher means and lower variances. Our analysis has been both short run and “local,” in the sense that the current behavior of the agent is taken as given. Future work could extend our local analysis to a “global” one. Such an extension should allow us to study the long run behavior of learning rules with differing risk attitudes.

From the characterization of monotonically mean-variance-averse rules we see that concerns for variance must pay some cost in terms of expected payoffs. This is because such rules cannot be in the class of rules studied by Börgers et al. Consequently, there must exist environments in which the concern for variance leads to no

improvement in terms of means. That a concern for variance impedes responsiveness to expected payoffs has been observed in experiments.

How exactly does learning respond to mean and variance? Existing research accounts for several psychological attitudes in the learning rule, though none seem to have explicitly taken into account the responsiveness of learning to mean *and* variance (see, e.g., Camerer and Ho [7]). Future work could attempt to more precisely determine how the learning rules people actually use respond to the mean and the variance of a distribution. Our characterization of monotonically mean-variance-averse learning rules should prove useful for this enterprise.

CHAPTER IV

MONOTONE IMITATION

A. Introduction

People often learn from their own experiences and by observing the experiences of others. As decision makers, we compare the performance of our decisions with the performance of the actions selected by other people. Actions leading to better results are more likely to be played in the future when the same problem is faced again. In this essay we find conditions, for this manner of learning, under which the number of individuals who play the actions with the best payoff distribution available is expected to increase. This requires the social learning process to exhibit three basic features. First, the action each individual plays in the next period is either the action she played in the current period or an action played by other individual she observed. Second, the higher the payoff provided by an action, the more likely this action is to be played in the next period. Third, all the actions are treated in a pair wise symmetric way.

We analyze of a finite population of individuals facing the same multi-armed bandit repeatedly. The information context we analyze corresponds to the one introduced by Schlag [33]: every period each member of the population (she) has to choose one action out of a fixed set that is common to the entire population. The decision makers do not know the payoff distributions of these actions; the only information available to each of them is the action that she played last period, the action played by some other member of the population (he), and the corresponding payoffs. This information determines the probability with which the individual will play each action in the next period. The function mapping the available information to the

probability of choosing each action is what we call an *imitation rule*, or, simply, the rule.

We study how the imitation rule used by the individuals determines whether a greater fraction of the population will, in the future, choose actions that are more likely to give higher payoffs. An imitation rule is said to be first-order monotone (FOM) if the fraction of the population who plays the actions whose distributions are first-order stochastically dominant is expected to increase. This is required for every possible set of probability distributions associated to the different actions.

The motivation for studying FOM rules is related to the way we interpret the payoffs of this model. In multi-armed bandits applied to economic analysis, typically payoffs are interpreted as units of money. For example, in Rothschild [30], payoffs represent the profits of a store trying to price an item in its inventory; and in the recent experiments of Apesteguia, Huck and Oechssler [1], payoffs correspond to the profits of oligopolistic firms playing a Cournot game. When payoffs are interpreted as monetary units, our focus on rules that lead to play actions with first-order stochastic dominant payoff distributions is consistent with the developments in the theory of rational choice under risk (Expected Utility Theory) with regards to stochastic dominance.¹ In Expected Utility Theory (EUT), decision makers with increasing Bernoulli utility functions prefer first-order stochastically dominant distributions. However, they do not always prefer payoff distributions with the highest expected value. An alternative approach would be focusing on imitation rules that lead to expected payoff maximizing actions. From the analysis we provide below, it is not difficult to see that such an approach would lead to virtually the same set of *improving* rules characterized by Schlag [33]. This is the set of rules for which the average payoff of the population is

¹See, for example, Hanoch and Levy [19].

expected to increase in every environment. Using Schlag's [33] characterization for improving rules and the characterization of FOM rules we provide here, it is easy to see that every (non trivial) improving rule is FOM. In EUT, the class of Bernoulli functions consistent with first-order stochastic dominance contains all the Bernoulli functions that are increasing, not only those that are linear. In the same way, FOM rules do not need to be linear in payoffs as the set of improving rules. In other words, our definition of FOM rules allows for a wider range of specifications because it is based on the notion of first-order stochastic dominance and not in expected values.

Our contribution is closely related to the line of research outlined in Börgers (1996). We relate a model of social learning to rational choice: we find the rules that lead a population facing repeatedly a multi-armed bandit to choose the actions that would be chosen by rational decision makers if the payoff distributions associated to the actions were known. Börgers (1996) suggests that a main role of learning and evolution theory in economics should be providing a theoretical framework to assess when experience will lead individuals to behave as rational agents. We derive the precise characteristics that a social learning process has to satisfy in order to lead the population to choose the actions which fully-informed rational agents would choose. Consistence of decisions with the order defined by first-order stochastic dominance may be regarded as one of the most elemental characteristics of rational choice. Furthermore, our results reveal that very simple and rather natural rules may display this property.²

Alternatively, payoffs may be (more generally) interpreted as a numerical representation of a preference order over outcomes associated to the played actions. A sufficient condition for this representation is that preferences over outcomes induce

²The concept of first-order monotonicity may also be motivated by evolutionary considerations. See the discussion in Section H.

a linear order over this set and that the cardinality of the set of outcomes is smaller than the cardinality of the set of payoffs in the analysis.³ Under this interpretation FOM rules are those leading individuals to choose actions that are more likely to provide preferred outcomes.

After introducing the framework in Section B, in Section C we provide two characterizations for FOM rules. The first characterization reveals that a necessary condition for first-order monotonicity is that the imitation rule must be *imitative*. An imitation rule is called imitative if each individual plays either the action that she played in the previous period or the action played by the individual she observed. Furthermore, this characterization shows that when the payoff distribution of one action strictly first-order stochastically dominates the payoff distribution of the other, then when two individuals playing each of these actions sample each other, the probability that the individual playing the dominated action switches to the dominant action is higher than the probability that the individual playing the dominant action switches to the dominated action. The analysis also reveals that this condition requires the imitation rule to be *impartial*. An imitation rule is called impartial if, when all the distributions of payoffs associated to the different actions are the same, the proportion of individuals who choose each action is expected to remain the same.

In the second characterization we describe FOM rules in terms of their functional form. In particular, besides being imitative, FOM rules can be described in terms of what we call the net-switching functions of an imitation rule. For each pair of actions a and a' , we define the *net-switching function from a to a'* . This is the difference between the probability of playing a' in the next period by an individual

³In other words, if it is possible to define an injective function from the set of outcomes to the set of payoffs. If a preference order is indifferent over two or more different outcomes, to obtain a linear order, the set of outcomes can be redefined in such a way that these outcomes are the same.

who played a in this period and observed another individual who played a' , and the probability of playing action a in the next period by an individual who played a' in this period and observed another individual who played a , when the payoff obtained from each action in both cases is the same. The arguments of this function are the payoffs obtained from playing each action. This characterization reveals that the net-switching functions are symmetric in the obtained payoffs in the sense that if the payoff obtained with action a is substituted for the payoff obtained with action a' and vice versa, then only the sign of the net-switching function changes. The magnitude remains the same. From the proof of this result it is easy to see that this feature is derived from the impartial property. Furthermore we prove that the net-switching function from a to a' is increasing in the payoff obtained with action a' . Likewise, and as a consequence of the symmetry described above, the net-switching function is decreasing in the payoff obtained with action a . Finally, we show that the net-switching functions are strictly positive if the payoff of the action to which the probability is being switched is strictly greater than the payoff of the action from which the probability is being switched. However, the net-switching functions do not need to be strictly increasing in the payoff of the action that receives the probability.

In Section D we study the dynamics of optimal choice in large populations. We show that if decision makers use a FOM rule, then it is very likely that an arbitrarily large fraction of the population will play an optimal action after a finite number of periods.

In Section E we discuss a number of examples in the literature that satisfy the properties we described above. It is easy to show that the rule Imitate if Better and Schlag's [33] Proportional Imitation Rule are FOM. As suggested by its name, the rule Imitate if Better prescribes switching (with probability one) to the observed action only if the payoff of that action is higher. Instead, the Proportional Imitation Rule

prescribes switching to the observed action only if the payoff of that action is higher and with a probability proportional to the difference in the payoffs. We also show that no FOM imitation rule can be said to be dominant in the sense that for all of them we can find a set of distributions for the payoffs such that, in that set, another FOM imitation rule has a greater expected increase in the fraction of the population playing an optimal action. Section F describes the analysis of properties founded in criteria related to performance at the individual level. Section G explains how to introduce risk aversion criteria in the analysis of social learning. Finally, Section H discusses alternative directions for further research.

Learning in social contexts has recently received a great deal of attention in economics. The work of Ellison and Fudenberg [14], [15] shows how simple imitation rules can lead a population to play the optimal action in the long run. Schlag [33] and Morales [25] proceed in the opposite direction. They propose a number of properties that are desirable for an imitation rule, in terms of their performance, and identify the imitation rules satisfying those properties. Our work is also in this direction, but in contrast to them, this essay focuses on properties of performance based on the concept of stochastic dominance rather than the expected value of the payoffs.

Another information context is considered in Schlag [33]. He analyzes a model similar to the one analyzed here, but the decision maker is able to observe the actions and obtained payoffs of two other individuals of the population, instead of only one. Schlag [34] identifies imitation rules that lead an infinite population to the expected payoff maximizing action. In Morales [25] the decision maker observes his action and obtained payoff and the action played by another member of the population and the corresponding payoff. He is able to identify the imitation rules that are absolutely expedient, i.e., imitation rules for which the expected payoff of each individual is expected to increase in every environment. Agents' information in Morales [25] is the

same as in the model discussed here. What is distinctive in his setup is the imitative restriction he imposes. This restriction requires that only the probabilities of the played and observed actions to be updated while the probabilities of playing all the other actions remain the same.

Other information settings have been studied in Vega-Redondo [37] and in the work of Apesteguia et al. [1], in the context of Cournot games. In Vega-Redondo [37], in every period, and with some positive probability, each firm can revise the action it has been playing. He assumes that in the next period firms would play any of the actions that obtained the highest profits in the current period and shows that the implied dynamics of the system lead firms to Walrasian behavior. Note that such decision rules imply that each firm is able to find out what are the quantities produced by the firms who obtained the highest profits in the periods when they revise their own production levels. Apesteguia et al. provide some experimental results that show how the imitation patterns of experimental subjects respond to different information treatments. An important result in their experiments, among many others, is that the probability of switching to observed actions responds to the magnitude of the difference in payoffs.

B. Framework

We analyze the behavior of a population described as a set of agents W . For most of our analysis the size of the population is finite, so $|W| < \infty$. In every period, each member of the population has to choose an action $a \in A$, where A is the finite set of actions available to the decision maker. The chosen action, a , yields a payoff $x \in [0, 1]$; this payoff is a random variable whose probability measure and distribution are denoted by μ_a and F_a , respectively. We will refer to the vector of distributions as

the *environment* and we will denote it by F , i.e. $F = (F_a)_{a \in A}$. We assume that the payoffs obtained from different actions are pair-wise independent and independent in time. The fraction of the population that chooses the action a in the current period is denoted by p_a . In what follows, first-order stochastic dominance is abbreviated by F_a fofd $F_{a'}$ and means that $F_a(x) \leq F_{a'}(x)$ for all $x \in [0, 1]$; and strict first-order stochastic dominance, denoted by F_a sfofd $F_{a'}$, means that F_a fofd $F_{a'}$ and $F_a(x) < F_{a'}(x)$ for at least one element $x \in [0, 1]$. Let $A^* := \{a \in A : F_a \text{ fofd } F_{a'} \forall a' \in A\}$ and for every set $S \subseteq A$, let $p(S)$ be the fraction of the population that chooses, in the current period, an action contained in this set, i.e. $p(S) := \sum_{a \in S} p_a$.

Each member of the population is able to observe the action that she chose in the current period and the payoff she obtains. She is also able to observe the action chosen in the current period by one of the other members of the population and the payoff he obtains. At the moment of choosing their actions in the next period this is all the information the individuals use. Decisions are assumed to be probabilistic, i.e., we assume that the action each decision maker chooses in the next period is random. The available information, however, affects the probability with which each action is chosen. In particular, we assume that the behavior of each individual in the population can be described by the function $L : A \times [0, 1] \times A \times [0, 1] \rightarrow \Delta(A)$. This function maps each quartet (a, x, a', y) to a vector $L(a, x, a', y)$. Here a is the action chosen by the agent in the current period, x is the payoff she obtains, a' is the action chosen by the agent that she observes and y is the payoff he obtains. All of these variables will determine the vector $L(a, x, a', y)$ containing the probabilities of choosing each action in the next period. The element $L(a, x, a', y)_{a''}$ of $L(a, x, a', y)$ denotes the probability with which the decision maker will play action a'' in the next period. $L(a, x, a', y)$ must be contained in $\Delta(A)$, which is the set of all probability distributions over A . Accordingly, the vector-valued function L is called the *imitation rule* of

the individual. For notational convenience, we assume that all the individuals in the population use the same rule. However, it is clear from the analysis that, with the only exception of the results provided in Section D, this assumption is not required. Note that if we know the imitation rule, the action chosen by the agent and the action chosen by the agent that she is able to observe, we can compute the expected probability of playing each action in the next period. Let $L_{a,a'}^{a''}$ be the expected probability, before the realization of the payoffs, of choosing action a'' tomorrow by a member of the population with imitation rule L who played action a and observed another individual who played action a' , i.e., $L_{a,a'}^{a''} := \int \int L(a, x, a', y)_{a''} dF_a(x) dF_{a'}(y)$.

For each member of the population, the individual that she is able to observe is determined by a random procedure. Formally, the probability that the agent $c \in W$ observes another agent $d \in W$ will be denoted by $\Pr(cd)$; thus $\sum_{d \in W \setminus \{c\}} \Pr(cd) = 1$ for all $c \in W$. It is assumed that the sampling process is symmetric, i.e., $\Pr(cd) = \Pr(dc)$ for all $c, d \in W$. Without this assumption the analysis rapidly becomes more complex. Furthermore, we assume that $\Pr(cd) > 0$ for all $c, d \in W$. As we will see in the next section, this assumption guarantees that any individual playing a non-optimal action can learn from any other individual playing an optimal action with some positive probability, and therefore, it allows the fraction of the population playing optimal actions to increase in expected terms.

The expected fraction of the population that will play action $a \in A$ in the next period, given the choices of the population in the current period, is denoted by p'_a . Let $s(c)$ be the action $c \in W$ played in the current period. It is easy to see that p'_a can be computed as

$$p'_a = \frac{1}{|W|} \sum_{c \in W} \sum_{d \in W \setminus \{c\}} \Pr(cd) L_{s(c), s(d)}^a. \quad (4.1)$$

Likewise, if we denote the expected fraction of the population that will play an action in A^* during the next period by $p'(A^*)$, it is easy to see that

$$p'(A^*) = \sum_{a \in A^*} \frac{1}{|W|} \sum_{c \in W} \sum_{d \in W \setminus \{c\}} \Pr(cd) L_{s(c), s(d)}^a. \quad (4.2)$$

The following definition formalizes one of the main concepts under analysis; it defines the imitation rules that lead a population to increase the number of its members that play an optimal action.

Definition 14 *A rule L is said to be first-order monotone (FOM) if $p'(A^*) \geq p(A^*)$, with strict inequality when $p(A^*) \in (0, 1)$, in every environment.*

Definition 14 does not impose extra restrictions on the random process that determines how individuals sample each other. Furthermore, it does not specify what the proportions of the population playing each action in the current period are. Therefore $p'(A^*) \geq p(A^*)$ is required to be satisfied for any environment, any symmetric sampling procedure where each individual may observe any of the others with a positive probability, and regardless the actions that are currently being played by the members of the population. Furthermore, in expected terms, the fraction of the population playing an optimal action always increases *strictly* as long as at least one individual plays an optimal action and at least one individual does not. Obviously $p(A^*)$ can not increase strictly when it is equal to one. The reason why we do not impose $p(A^*)$ to increase strictly when $p(A^*) = 0$ is a little more subtle and we explain it in the next section.

C. First-Order Monotone Imitation Rules

In this section we provide two characterizations for FOM rules. The first characterization reveals that these rules need to satisfy a number of restrictions. The first

restriction requires that each member of the population will not choose in the next period any action that is not either the action she chose or the one that was chosen by the agent she observed. Schlag [33] calls such rules *imitative*.

Definition 15 *A rule L is said to be imitative if for all actions $a, a', a'' \in A$, such that $a'' \notin \{a, a'\}$, we have $L(a, x, a', y)_{a''} = 0$ for all $x, y \in [0, 1]$.*

The following preliminary result describes the expected change in the fraction of the population who plays action $a \in A$ for imitative rules.

Lemma 6 *If L is imitative, then*

$$p'_a - p_a = \frac{1}{|W|} \sum_{c \in W, s(c) \neq a} \sum_{d \in W, s(d)=a} \Pr(cd) (L_{s(c),a}^a - L_{a,s(c)}^{s(c)}). \quad (4.3)$$

The expression above is just the difference between the expected fraction of the population that will switch from an action $a' \in A \setminus \{a\}$ to action a in the next period and the expected fraction of the population that will switch from action a to a different action in the next period. The proof of Lemma 6 follows from straightforward calculations and the assumption that $\Pr(cd) = \Pr(dc)$ for all $c, d \in W$.

We also found that every FOM imitation rule is not expected to change the proportion of the population who chooses each action in all the environments where the distributions of payoffs associated with each action are the same. We call that property *impartiality*.

Definition 16 *We say that a rule is impartial if $p'_a - p_a = 0$ for all $a \in A$, whenever $F_a(x) = F_{a'}(x)$ for all $x \in [0, 1]$ for all $a' \in A$.*

We now present a first characterization of FOM rules. This characterization shows that both imitation and impartiality are necessary conditions for a rule to be FOM.

Lemma 7 *A rule L is FOM if and only if it satisfies the following conditions:*

(i) *L is imitative.*

(ii) *$F_{a'} \text{ sfosd } F_a \Rightarrow L_{aa'}^{a'} - L_{a'a}^a > 0, \forall a, a' \in A$ for all environments.*

Proof.

Sufficiency:

Since L is imitative, Lemma 0 applies. Therefore, for all $a \in A$ we have

$$p'_a - p_a = \frac{1}{|W|} \sum_{c,d \in W; s(c) \neq a, s(d)=a} \Pr(cd)(L_{s(c),a}^a - L_{a,s(c)}^{s(c)}). \quad (4.4)$$

Now, because of (ii), for every $a \in A^*$ and $c \in W$, we have $L_{s(c),a}^a - L_{a,s(c)}^{s(c)} \geq 0$. It follows that $p'_a - p_a \geq 0$ for all $a \in A^*$, therefore $p'(A^*) \geq p(A^*)$. Note that when $p(A^*) \in \{0, 1\}$ we have that $p'(A^*) = p(A^*)$, because L is imitative. Now consider $p(A^*) \in (0, 1)$. We know that $p'_{a'} - p_{a'} \geq 0$ for all $a' \in A^*$. Furthermore, by Lemma 6, for some $a' \in A^*$, we have $p'_{a'} - p_{a'} > 0$, because $p(A \setminus A^*) > 0$, $\Pr(cd) > 0$ for all $c, d \in W$, and $L_{a,a'}^{a'} - L_{a',a}^a > 0$ for all $a \in A \setminus A^*$.

Necessity of (i):

Suppose that there are some $x, y \in [0, 1]$, $a, a', a'' \in A$, $a'' \notin \{a, a'\}$ such that $L(a, x, a', y)_{a''} > 0$. Furthermore, suppose $p_a + p_{a'} = 1$, $F_a = F_{a'}$, $\mu_a(x) = \mu_a(y) = \mu_a(1) = 1/3$, and $\mu_{a''}(0) = 1$ for all $a''' \in A \setminus \{a, a'\}$.⁴ It follows that $A^* = \{a, a'\}$, therefore $p(A^*) = 1$. Suppose that there are $c, d \in W$ such that $\Pr(cd) > 0$, $s(c) = a$, and $s(d) = a'$, then $p'(A^*) < 1$. Thus, L has to be imitative.

Necessity of condition (ii):

Suppose that $F_{a'} \text{ sfosd } F_a$, but $L_{aa'}^{a'} - L_{a'a}^a \leq 0$, for some $a, a' \in A$. Suppose that $p_a + p_{a'} = 1$, $p_a < 1$, $p_{a'} < 1$ and that $A^* = \{a'\}$. Since L is imitative, Lemma 6

⁴If $x \neq y = 1$, then it means that $\mu_a(1) = 2/3$ and $\mu_a(x) = 1/3$; if $x = y = 1$, then $\mu_a(1) = 1$; etc. In what follows we adopt this notation convention.

yields

$$p'_{a'} - p_{a'} = \frac{1}{|W|} \sum_{c,d \in W; s(c)=a, s(d)=a'} \Pr(cd)(L'_{a,a'} - L^a_{a',a}). \quad (4.5)$$

Assume that $s(c) = a$ and $s(d) = a'$ for some $c, d \in W$. Then, we have that $p(A^*) \in (0, 1)$, but $L'_{aa'} - L^a_{a'a} \leq 0$ implies $p'(A^*) \leq p(A^*)$. \square

Corollary 3 *If L is FOM, then $F_{a'}$ fofd F_a $L'_{aa'} - L^a_{a'a} \geq 0$, $\forall a, a' \in A$, for all environments.*

Proof. Suppose that for some $a, a' \in A$, $F_{a'} = F_a$, but $L'_{aa'} - L^a_{a'a} < 0$. For $\varepsilon \in (0, 1)$ consider the modified environment \tilde{F} such that for any interval $I \subseteq [0, 1)$, $\tilde{\mu}_{a'}(I) = (1 - \varepsilon)\mu_{a'}(I)$, $\tilde{\mu}_{a'}(1) = \mu_{a'}(1) + \varepsilon\mu_{a'}[0, 1)$; and for any interval $I \subseteq (0, 1]$, $\tilde{\mu}_a(I) = (1 - \varepsilon)\mu_a(I)$, $\tilde{\mu}_a(0) = \mu_a(0) + \varepsilon\mu_a(0, 1]$. In this modified environment $\tilde{F}_{a'}$ sfosd \tilde{F}_a . Since $L'_{a,a'} - L^a_{a',a} < 0$ in the original environment, and $\tilde{L}'_{a,a'} - \tilde{L}^a_{a',a}$ can be written as a continuous function in ε , we obtain that for small enough ε , $\tilde{L}'_{aa'} - \tilde{L}^a_{a'a} < 0$. This is a contradiction because $\tilde{F}_{a'}$ sfosd \tilde{F}_a . \square

Remark 5 *Note that Corollary 3 implies that if $F_a(x) = F_{a'}(x)$ for all $x \in [0, 1]$, then $L'_{aa'} - L^a_{a'a} = 0$. From Lemma 6, this implies that every FOM rule is impartial.*

As revealed by the proof of Lemma 7, the imitative property of FOM rules follows from the fact that improvement is required in every environment, regardless of the fraction of the population that is playing each action in the current period. In particular, if all the members of the population play an action in A^* , then the expected fraction of the population that plays an action in A^* in the next period will be strictly less than one unless all of them, with probability one, play either the action they played in the current period or the action played by the other member of the population they observed. This requires the imitation rule to be imitative. This condition implies that when no individual in the population is playing an optimal

action the population is unable to learn to play actions in A^* . For this reason we do not impose that $p'(A^*) > p(A^*)$ when $p(A^*) = 0$. Clearly an imitation rule that satisfies $p'(A^*) > p(A^*)$ when $p(A^*) = 0$ and $p'(A^*) = p(A^*)$ when $p(A^*) = 1$ can not exist.⁵

The necessity of properties that are similar to our notion of impartiality has been found in many places in the literature. From Schlag's [33] results it follows that, if the average payoff of the population is expected to increase in every environment, then the imitation rule of the population must verify that, in every environment where all the actions have the same expected payoffs, the expected fraction of the population who plays each action in the next period is the same as today. Similar results for individual learning have been found by Börgers et al. and in Chapter II.

The next result provides a characterization of FOM social rules in terms of the shape of $L(a, x, a', y)_{a''}$, for all $a, a', a'' \in A$. In the analysis below, we will recurrently use the concept of *net-switching function* from action a to action a' . We denote this function by $g_{aa'}(x, y)$ and it is defined as $g_{aa'}(x, y) \equiv L(a, x, a', y)_{a'} - L(a', y, a, x)_a$. This function measures how much probability is being shifted from action a to action a' when two agents playing a and a' observe each other.

Proposition 9 *A rule L is FOM if and only if it satisfies the following conditions:*

- (i) *L is imitative,*
- (ii) *for all $a, a' \in A$, the net-switching functions $g_{aa'}(x, y)$ satisfy:*
 - (ii.1) $g_{aa'}(x, y) = -g_{aa'}(y, x), \forall x, y \in [0, 1],$
 - (ii.2) $g_{aa'}(x, y)$ *is non-decreasing in y , for all $x \in [0, 1].$*

⁵Alternatively, for every FOM rule we could allow each individual to experiment with a small probability. This would imply that $p'(A^*) > p(A^*)$ when $p(A^*) = 0$. However, then we would have $p'(A^*) < 1$ when $p(A^*) = 1$. Nevertheless, it is clear that, for any environment, one could have an small enough experimentation probability such that $p'(A^*) > p(A^*)$ when $p(A^*) \in (0, 1)$.

(ii.3) $g_{aa'}(x, y) > 0, \forall x, y \in [0, 1]$ such that $y > x$.

In order to prove Proposition 9, we need the following lemma.

Lemma 8 *Consider a function $g : [0, 1]^2 \rightarrow [-1, 1]$ that satisfies:*

- (i) $g(x, y) = -g(y, x), \forall x, y \in [0, 1]$,
- (ii) $g(x, y)$ is non-decreasing in y , for all $x \in [0, 1]$,
- (iii) $g(x, y) > 0, \forall x, y \in [0, 1]$ such that $y > x$.

Then, for any two independent random variables X and Y , such that Y sfosd X , we have that $E(g(X, Y)) > 0$.

Proof. The proof proceeds in two steps:

Step 1: We first prove that Y sfosd X implies $P(Y > X) > P(Y < X)$. Let $M := \{z \in [0, 1] : F_X(z) > F_Y(z)\}$ denote the set of all points in $[0, 1]$ where the distribution function of X is strictly larger than the distribution function of Y . M is non-empty since Y sfosd X . We show $P_X(M) > 0$, where P_X is the measure induced by X . If $0 \in M$ then $P_X(M) \geq F_X(0) > 0$. So, without loss of generality we will assume $0 \notin M$. Let $\tilde{z} \in M$. Define $z_0 := \sup\{z : z \leq \tilde{z}, F_X(z) = F_Y(z)\} \geq 0$. If $z_0 \in M$, then $\lim_{z \rightarrow z_0-} F_X(z) = \lim_{z \rightarrow z_0-} F_Y(z)$ because of the definition of z_0 and

$$\begin{aligned}
 P_X(M) &\geq P(X = z_0) \\
 &= F_X(z_0) - \lim_{z \rightarrow z_0-} F_X(z) \\
 &= F_X(z_0) - \lim_{z \rightarrow z_0-} F_Y(z) \\
 &\geq F_X(z_0) - F_Y(z_0) \\
 &> 0.
 \end{aligned} \tag{4.6}$$

If $z_0 \notin M$, then $F_X(z) > F_Y(z)$ for all $z \in (z_0, \tilde{z}] =: N$ and

$$\begin{aligned}
P_X(M) &\geq P_X(N) \\
&= F_X(\tilde{z}) - F_X(z_0) \\
&= F_X(\tilde{z}) - F_Y(z_0) \\
&\geq F_X(\tilde{z}) - F_Y(\tilde{z}) \\
&> 0.
\end{aligned} \tag{4.7}$$

We define now $h(a, b) := 1_{\{b > a\}}$ and $\Delta h(a, b) := h(a, b) - h(b, a)$. For all $a \in M$, we have the inequality

$$\begin{aligned}
E(\Delta h(a, Y)) &= 1 - F_Y(a) - F_Y(a-) \\
&> 1 - F_X(a) - F_X(a-) \\
&= E(\Delta h(a, X)).
\end{aligned} \tag{4.8}$$

The inequality holds weakly for $a \notin M$. Let $\tilde{X} \stackrel{d}{=} X$ be a random variable which is independent from X and Y . Then we get

$$\begin{aligned}
E(\Delta h(X, Y)) &= E(E(\Delta h(X, Y)|X)) \\
&> E(E(\Delta h(X, \tilde{X})|X)) \\
&= E(\Delta h(X, \tilde{X})) \\
&= 0
\end{aligned} \tag{4.9}$$

and therefore, using the definition of Δh , we obtain $P(Y > X) > P(Y < X)$.

Step 2: Because of the monotonicity of the probability measure P , there exists $\epsilon > 0$ such that $P(Y - X > \epsilon) > P(X > Y)$. Let $D := \{(x, y) : x, y \in [0, 1], y - x > \epsilon\}$.

Now, we prove that

$$c := \inf\{g(x, y) : x, y \in D\} > 0. \tag{4.10}$$

Let $k := \inf\{n \in \mathbb{N} : n > 2/\epsilon\}$ and $I_i := [0, \frac{i-1}{k}] \times [\frac{i}{k}, 1]$ for all $i \in \{2, \dots, k-1\}$. It is easy to see that $D \subset \bigcup_{i=2}^{k-1} I_i$. Then $c_i := \inf\{g(x, y) : x, y \in I_i\} = g(\frac{i-1}{k}, \frac{i}{k}) > 0$ for all $i \in \{2, \dots, k-1\}$ because of the monotonicity of g and assumption (iii). Then $c \geq \min\{c_i : i \in \{2, \dots, k-1\}\} > 0$.

Let

$$g_1(x, y) := (g(x, y) - c) \cdot 1_{\{g(x, y) \geq c\}} + (g(x, y) + c) \cdot 1_{\{g(x, y) \leq -c\}}, \quad (4.11)$$

and

$$g_2(x, y) := c \cdot (1_{\{g(x, y) \geq c\}} - 1_{\{g(x, y) \leq -c\}}) + g(x, y) \cdot 1_{\{-c < g(x, y) < c\}}. \quad (4.12)$$

We have $g(x, y) = g_1(x, y) + g_2(x, y)$ where g_1, g_2 satisfy conditions (i) and (ii) of the hypothesis. It follows that $E(g_1(X, Y)) \geq 0$ because

$$\begin{aligned} E(g_1(X, Y)) &= \int \int g_1(x, y) dF_Y(y) dF_X(x) \\ &\geq \int \int g_1(x, y) dF_X(y) dF_X(x) \\ &\geq \int \int g_1(x, y) dF_Y(x) dF_X(y) \\ &= - \int \int g_1(y, x) dF_Y(x) dF_X(y) \\ &= -E(g_1(X, Y)), \end{aligned} \quad (4.13)$$

where we have used the independence of X and Y and the fact that $g_1(x, y)$ is non-decreasing in the second argument and non-increasing in the first argument. On D , we have $g_2(x, y) = c$ and on $C := \{(x, y) : x > y\}$, we have $g_2(x, y) \geq -c$. Thus, we get

$$E(g_2(X, Y)) \geq c \cdot P(D) - c \cdot P(C) > 0. \quad (4.14)$$

Summarized, we have $E(g(X, Y)) > 0$. \square

Proof of Proposition 9.

Necessity of (ii.1):

We start by proving that $g_{aa'}(x, x) = 0$ for all $x \in [0, 1]$ and all $a, a' \in A$. Consider an arbitrary $x \in [0, 1]$ and an environment where $\mu_a(x) = \mu_{a'}(x) = 1$. Notice that $F_a = F_{a'}$. Therefore, by Remark 5, $L_{aa'}^{a'} - L_{a'a}^a = 0$. But in this environment $L_{aa'}^{a'} - L_{a'a}^a = g_{aa'}(x, x)$. It follows that $g_{aa'}(x, x) = 0$.

Now, for any $x, y \in [0, 1]$ and some $a, a' \in A$, consider an environment where $\mu_a(x) = \mu_{a'}(x) = \mu_a(y) = \mu_{a'}(y) = 1/2$, and notice that $F_a = F_{a'}$. By Remark 5,

$$\begin{aligned} 0 &= L_{aa'}^{a'} - L_{a'a}^a \\ &= \frac{1}{4}(g_{aa'}(x, y) + g_{aa'}(y, x) + g_{aa'}(x, x) + g_{aa'}(y, y)) \\ &= \frac{1}{4}(g_{aa'}(x, y) + g_{aa'}(y, x)). \end{aligned} \tag{4.15}$$

Therefore, $g_{aa'}(x, y) = -g_{aa'}(y, x)$.

Necessity of (ii.2):

Suppose that for some $a, a' \in A$ and $x, y', y \in [0, 1]$ such that $y' > y$, we have that $g_{aa'}(x, y') < g_{aa'}(x, y)$. Consider an environment such that $\mu_a(x) = \mu_{a'}(x) = 1 - \varepsilon$, $\mu_a(y) = \mu_{a'}(y') = \varepsilon$, and notice that $F_{a'} = F_a$. It follows that

$$\begin{aligned} 0 &< L_{aa'}^{a'} - L_{a'a}^a \\ &= (1 - \varepsilon)\varepsilon g_{aa'}(x, y') + (1 - \varepsilon)\varepsilon g_{aa'}(y, x) + (1 - \varepsilon)^2 g_{aa'}(x, x) + \varepsilon^2 g_{aa'}(y, y') \\ &= \varepsilon[(1 - \varepsilon)\{g_{aa'}(x, y') - g_{aa'}(x, y)\} + \varepsilon g_{aa'}(y, y')]. \end{aligned} \tag{4.16}$$

The last equality holds because of the necessity of (ii.1). The term inside the brackets $\{.\}$ is negative. Thus, for small enough ε , the term inside the brackets $[.]$ is negative, causing a contradiction.

Necessity of (ii.3):

Suppose that $y > x$ and $g_{aa'}(x, y) \leq 0$. Consider $a, a' \in A$ such that $\mu_a(x) = \mu_{a'}(y) = 1$. Clearly $F_{a'}$ sfosd F_a , but $L_{aa'}^{a'} - L_{a'a}^a = g_{aa'}(x, y) \leq 0$, which violates condition (ii) of Lemma 7.

Sufficiency:

Sufficiency follows directly from Lemma 8. \square

Proposition 9 reveals the properties of the functional form of FOM imitation rules that we described in the Introduction. Besides being imitative, these rules treat every pair of actions in a pair wise symmetric way (condition (ii.1)). Furthermore, net-switching from action a to action a' must be a non-decreasing function of the payoff of action a' (condition (ii.2)), and strictly positive if the payoff of a' is greater than the payoff of a (condition (ii.3)). This result allows us to check, directly from the functional form of an imitation rule, whether it is FOM or not. In Section E we analyze a number of imitation rules from the literature and use Proposition 9 to assess whether they are FOM.

The following corollary follows directly from conditions (ii.1) and (ii.2) in Proposition 9.

Corollary 4 *If L is FOM then $g_{aa'}(x, y)$ is non-increasing in x , for all $y \in [0, 1]$.*

In our definition of A^* , all the distributions in this set are the same. This can be slightly generalized. If A can be partitioned in A^{**} and $A \setminus A^{**}$ in such a way that for all actions $a \in A^{**}$ and $a' \in A \setminus A^{**}$, we have that F_a sfosd $F_{a'}$, then the set of rules that satisfy $p'(A^{**}) \geq p(A^{**})$, with strict inequality when $p(A^*) \in (0, 1)$, in every environment, is equivalent to the set of FOM rules.⁶

⁶Clearly every rule that satisfies $p'(A^{**}) \geq p(A^{**})$ in every environment, with strict inequality in when $p(A^{**}) \in (0, 1)$ has to be FOM. The fact that FOM rules satisfy $p'(A^{**}) \geq p(A^{**})$ in every environment, with strict inequality when $p(A^{**}) \in (0, 1)$, follows directly from Lemma 6, Lemma 7, and straightforward algebraic manipulations.

D. Large Population Dynamics

In this section we analyze the dynamics of the fraction of the population playing an optimal action in large populations. We show that, with a probability arbitrarily close to one, the fraction of the population that plays an optimal action is arbitrarily close to one after a finite number of periods, provided that the population is large enough and decision makers use a FOM rule.

In the rest of this section we will assume that $\Pr(cd)$ does not depend on c, d for all $c, d \in W$. We will also assume that the sampling process is independent⁷. We use these assumptions below, when we approximate the behavior of a finite population by analyzing the dynamics of a population that is a continuum.

First, we introduce the dynamic system

$$p_a^{t+1} = \sum_{a' \in A} \sum_{a'' \in A} p_{a'}^t p_{a''}^t L_{a'a''}^a, \quad \forall a \in A, t = 0, 1, \dots$$

with $p_a^0 \in [0, 1]$ given $\forall a \in A$ and $\sum_{a \in A} p_a^0 = 1$. This system is called the *dynamics of infinite populations* because this expression can be interpreted as the motion equation of the fraction of the population that plays action a in a population that is a continuum: for each action $a' \in A$, the fraction of the population that plays action a' in this period and a in the next period is given by $p_{a'}^t \sum_{a'' \in A} p_{a''}^t L_{a'a''}^a$. Adding across actions $a' \in A$, we obtain that the total fraction of the population that will play action a in the next period, p_a^{t+1} , corresponds to the expression above. The fraction of the population that plays an action in A^* at time t will be denoted by $p^t(A^*)$. In other words, $p^t(A^*) := \sum_{a \in A^*} p_a^t$.

⁷Independent sampling means that the joint probability of the events c samples d and d samples c is given by $\Pr(cd)\Pr(dc)$ for all $c, d \in W$ such that $c \neq d$.

Remark 6 *If L is imitative, then the dynamics of infinite populations can be simplified to*

$$p_a^{t+1} = p_a^t + p_a^t \sum_{a' \in A} p_{a'}^t (L_{a'a}^a - L_{aa'}^{a'}), \quad \forall a \in A, t = 0, 1, \dots \quad (4.17)$$

The next result shows that, if the imitation rule used by the population is FOM, then, $p^t(A^*)$ converges to one for the dynamics of infinite populations.

Lemma 9 *Let L be a FOM rule and assume that $p^0(A^*) > 0$. Then the dynamics of infinite populations satisfy $\lim_{t \rightarrow \infty} p^t(A^*) = 1$.*

Proof. Suppose that $p^0(A^*) > 0$. Since L is FOM, the sequence $\{p^t(A^*)\}_{t=1}^\infty$ is monotonically increasing and bounded by 1. Thus, it is sufficient to prove that 1 is the least upper bound of this sequence. Suppose that the least upper bound is $p^* < 1$. Let

$$\gamma := \min_{a \in A^*, a' \in A \setminus A^*} \{L_{a'a}^a - L_{aa'}^{a'}\} \cdot (1 - p^*) > 0. \quad (4.18)$$

Then, we obtain that

$$\begin{aligned} p^{t+1}(A^*) &= \sum_{a \in A^*} (p_a^t + p_a^t \sum_{a' \in A \setminus A^*} p_{a'}^t (L_{a'a}^a - L_{aa'}^{a'})) \\ &\geq p^t(A^*) + \sum_{a \in A^*} p_a^0 \gamma. \end{aligned} \quad (4.19)$$

Since $\sum_{a \in A^*} p_a^0 \gamma > 0$, this causes a contradiction. \square

In our next result, we use Theorem 3 in Schlag [33] and Lemma 9 to prove that, with a probability arbitrarily close to one, the fraction of the population that plays an optimal action is arbitrarily close to one after a finite number of periods, provided that the population is large enough and the decision makers use a FOM rule. Let $p_a^{t,N}$ denote the fraction of the population that, at time t , plays a in a population of size N and let $p^{t,N}(A^*) := \sum_{a \in A^*} p_a^{t,N}$. First, we provide Theorem 3 in Schlag [33].

Theorem 1 (*Schlag [33]*) *Assume that the sampling process is independent and that $\Pr(cd)$ does not depend on c or d for all $c \neq d \in W$. Assume that all individuals use*

the rule L and $p_a^{0,N} = p_a^0$ for all $a \in A$. Then, for every $\varepsilon, \delta > 0$, $T \in \{1, 2, \dots\}$, and $(p_a^{0,N})_{a \in A}$, there exists $N_0 \in \{1, 2, \dots\}$ such that for any population of size $N > N_0$, the event $\{[\sum_{a \in A} (p_a^{T,N} - p_a^T)^2]^{1/2} > \delta\}$ occurs with probability less than ε .

Proposition 10 *If a rule L is FOM, and $p^{0,N}(A^*) > 0$, then for all $\varepsilon, \delta > 0$, there exist $N_0, T < \infty$ such that for all $N > N_0$, $P(1 - p^{T,N}(A^*) > \delta) < \varepsilon$.*

Proof. By Lemma 9, we know that there exists a natural number T such that $1 - p^T(A^*) < \delta/2$. Now, by Theorem 3 in Schlag [33], for this T there exists N_0 such that for all $N > N_0$ we have

$$P(|p^T(A^*) - p^{T,N}(A^*)| > \delta/2) < \varepsilon. \quad (4.20)$$

Then, by the triangle inequality, we obtain that $P(1 - p^{T,N}(A^*) > \delta) < \varepsilon$. \square

E. Examples

In this section we provide a number of examples of imitation rules that satisfy the properties studied above. We also show that no FOM rule can be said to be dominant in the sense of having the greatest $p'(A^*)$ in every environment among all the FOM rules.

Our first example is Schlag's [33] Proportional Imitation Rule. This is an imitative rule that never switches unless the payoff of the sampled action is strictly greater than the payoff of the action played by the individual. If this is the case, the probability of switching is $y - x$. This rule is defined as follows: for all $a, a' \in A$ let

$$L(a, x, a', y)_{a'} = \begin{cases} y - x & y > x \\ 0 & \text{otherwise,} \end{cases} \quad (4.21)$$

$$L(a, x, a', y)_a = \begin{cases} 1 - y + x & y > x \\ 1 & \text{otherwise.} \end{cases} \quad (4.22)$$

Straightforward calculations show that $g_{aa'}(x, y) = y - x$. It follows that this rule is FOM. Schlag [33] shows that this rule has a number of interesting properties. For example, it has the greatest increase in the expected average payoff of the population among all the imitation rules that are improving.

Our second example is a version of the rule *Imitate if Better*. This rule is an imitative rule that prescribes switching with probability one when $y > x$, with probability $1/2$ if $y = x$, and not to switch otherwise. This rule can be written as

$$L(a, x, a', y)_{a'} = \begin{cases} 1 & y > x \\ \frac{1}{2} & y = x \\ 0 & \text{otherwise} \end{cases} \quad (4.23)$$

$$L(a, x, a', y)_a = \begin{cases} 0 & y > x \\ \frac{1}{2} & y = x \\ 1 & \text{otherwise} \end{cases} \quad (4.24)$$

for all $a, a' \in A$.

For this rule

$$g_{aa'}(x, y) = \begin{cases} 1 & y > x \\ 0 & y = x \\ -1 & \text{otherwise} \end{cases} \quad (4.25)$$

for all $a, a' \in A$.

Therefore, this rule is FOM. As shown by Schlag [33], *Imitate if Better* is not improving, so it is not expected to increase the expected average payoff of the population in every environment. Can we claim that any of these rules is better than the other in any sense? Or more important, can we find any rule that is the “best”?

The most natural notion to assess the performance of different rules is based on the expected increase in the number of individuals that play the optimal action in the next period. Therefore, an action may be said to be the best if the expected increase in the fraction of the population that plays an optimal action in the next period is bigger than for any other rule, in every environment, regardless the initial state and sampling process. This is captured in the following definition.

Definition 17 *A FOM rule L is dominant if $p'_L(A^*) - p(A^*) \geq p'_{L'}(A^*) - p(A^*)$ for every FOM rule L' in every environment.*

The next result shows that such an imitation rule can not exist.

Proposition 11 *FOM dominant rules do not exist.*

Proof. The proof proceeds in two steps. First, we show that if a dominant rule exists, it has to be the rule Imitate if Better, which we denote *LIB*. This rule satisfies

$$g_{aa'}(x, y) = \begin{cases} 1 & y > x \\ 0 & y = x \\ -1 & \text{otherwise} \end{cases} \quad (4.26)$$

for all $x, y \in [0, 1]$, and $a, a' \in A$.⁸ Consider any other FOM rule L' such that $g'_{aa'}(x, y) < 1$ for some $y > x$. Consider an environment such that $\mu_{a'}(y) = \mu_a(x) = 1$. Assume $p_a + p_{a'} = 1$, $A^* = \{a'\}$ and $\Pr(cd) > 0$ for some $c, d \in W$, such that $s(c) = a$

⁸By Imitate if Better learning rules we mean any learning rule whose net switching functions $g_{aa'}(x, y)$ are like the ones specified above. In particular, $L(a, x, a', x)_{a'}$ does not need to be $1/2$ as in our earlier example.

and $s(d) = a'$. Then,

$$\begin{aligned}
p'_{L'}(A^*) - p(A^*) &= p'_{a'} - p_{a'} \\
&= \frac{1}{|W|} \sum_{c,d \in W; s(c)=a, s(d)=a'} \Pr(cd) (L'_{aa'} - L^a_{a'a}) \\
&= \frac{1}{|W|} \sum_{c,d \in W; s(c)=a, s(d)=a'} \Pr(cd) g'_{aa'}(x, y) \\
&< \frac{1}{|W|} \sum_{c,d \in W; s(c)=a, s(d)=a'} \Pr(cd) \\
&= p'_{LIB}(A^*) - p(A^*).
\end{aligned} \tag{4.27}$$

It follows that if a dominant rule exists, it has to be the rule Imitate if Better. Now, we prove that this rule is not dominant either. Consider the environment such that $\mu_{a'}(.05) = \mu_a(.05) = \mu_{a'}(.9) = \mu_a(.1) = 1/2$, and consider a FOM rule L' such that

$$g_{aa'}(x, y) = \begin{cases} 1 & y \geq .8 \wedge x \leq .2 \\ -1 & x \geq .8 \wedge y \leq .2 \\ \frac{1}{2}(1_{\{y>x\}} - 1_{\{x>y\}}) & \text{otherwise.} \end{cases} \tag{4.28}$$

It is easy to compute that for this rule, $L'_{a,a'} - L^a_{a',a} = 3/8$, while $LIB_{a,a'}^{a'} - LIB_{a',a}^a = 2/8$. As above, assume that $p_a + p_{a'} = 1$, $A^* = \{a'\}$, and that $\Pr(cd) > 0$ for some $c, d \in W$, such that $s(c) = a$ and $s(d) = a'$. With similar calculations to those in the first step, we obtain

$$p'_{L'}(A^*) - p(A^*) > p'_{LIB}(A^*) - p(A^*). \tag{4.29}$$

Thus, the rule Imitate if Better is not dominant either. \square

F. Individual Monotonicity

The motivation for the construction of FOM rules is based on the performance of these rules in terms of the fraction of the population that is expected to play an optimal action in the next period. The analysis provided so far has not considered

how the way in which each individual uses the information she gets helps her improve her own payoffs. Can the construction of FOM rules be derived from the analysis of individual performance? A first observation is that FOM rules are imitative. This feature follows from fundamental characteristics of the social learning process and the properties, in terms of social performance, to be obtained: imitation is required to avoid that a population already playing optimal actions to experiment non-optimal choices. In the analysis of individual choices, such an argument does not apply. Each individual must be exposed to go from optimal to non-optimal choices when learning takes place. As a consequence, when studying improved performance properties at the individual level, imitation can not be derived from the analysis as we did in the case of improved performance at the social level. Therefore, in order to derive FOM rules from properties related to individual level performance, imitation must be *assumed* –not *derived*. In this sense, an individual level derivation of FOM rules takes imitative behavior as given, but provides the specific ways in which imitation has to be implemented in order to satisfy improved performance at the individual level.⁹ Formally, we can characterize a family of imitative rules such that, if the payoff distribution of one of the observed actions first-order stochastically dominates the distribution associated with the other observed action, then the expected probability of playing the dominant action is higher than the expected probability of playing the dominated action. We call such rules *individually monotone* (IM). The characterization of these rules reveals that, indeed, they are a refinement of the family of FOM rules. Now we provide the formal analysis of IM rules

⁹Indeed, imitative behavior is sometimes considered as an intrinsic characteristic of the human decision process. For example, see Cubitt and Sugden [9].

Definition 18 *An imitative rule is said to be individually monotone if $F_{a'}$ sfosd F_a implies*

- (i) $L_{aa'}^{a'} > L_{aa'}^a$
- (ii) $L_{a'a}^{a'} > L_{a'a}^a$.

We can characterize individually monotone rules using the mathematical structures we discovered in the analysis of net-switching functions for FOM rules. Specifically, we find that the probability of switching, $L(a, x, a', y)_{a'}$, must be increasing in the payoff of the sampled action and symmetric with respect to payoffs, in the sense that, if payoffs of the actions are interchanged, the probabilities of playing them are interchanged too.¹⁰

Proposition 12 *An imitative rule L is individually monotone if and only if, for all $a, a' \in A$, $L(a, x, a', y)_{a'}$ satisfies:*

- (i) $L(a, x, a', y)_{a'} = L(a, y, a', x)_a, \forall x, y \in [0, 1]$,
- (ii) $L(a, x, a', y)_{a'}$ is non-decreasing in y , for all $x \in [0, 1]$,
- (iii) $L(a, x, a', y)_{a'} > 1/2$ for all $x, y \in [0, 1]$ such that $y > x$.

Proof. Suppose $F_{a'}$ fofd F_a . We need to verify

$$\int \int L(a, x, a', y)_{a'} dF_a(x) dF_{a'}(y) > \int \int L(a, x, a', y)_a dF_a(x) dF_{a'}(y). \quad (4.30)$$

But, since L is imitative, this is equivalent to

$$\int \int L(a, x, a', y)_{a'} dF_a(x) dF_{a'}(y) > 1/2. \quad (4.31)$$

¹⁰It may seem more attractive to find rules L such that, when $F_{a'}$ sfosd F_a , the lower bound for $L_{aa'}^{a'}$ is greater than $1/2$ for any environment. However, an argument similar to the one in the proof of Corollary 3 shows that such rules do not exist.

Let $h_{aa'}(x, y) := 2L(a, x, a', y)_{a'} - 1$. Then, the last expression is equivalent to

$$\int \int h_{aa'}(x, y) dF_a(x) dF_{a'}(y) > 0. \quad (4.32)$$

The arguments in the proof of Proposition 9 can be used to prove that a necessary and sufficient condition for the last statement to be true is that $h_{aa'}(x, y)$ satisfies

- (i) $h_{aa'}(x, y) = -h_{aa'}(y, x)$, $\forall x, y \in [0, 1]$,
- (ii) $h_{aa'}(x, y)$ is non-decreasing in y , for all $x \in [0, 1]$,
- (iii) $h_{aa'}(x, y) > 0$ for all $x, y \in [0, 1]$ such that $y > x$.

These conditions are equivalent to conditions (i)-(iii) in the hypothesis. \square

For example, the rule Imitate if Better discussed in Section E is IM if and only if $L(a, x, a', x) = 1/2$ for all $a, a' \in A$ and for all $x \in [0, 1]$. The Proportional Imitation Rule is not IM, but the rule given by

$$L(a, x, a', y)_a = \frac{1}{2}(1 + x - y) \quad (4.33)$$

$$L(a, x, a', y)_{a'} = \frac{1}{2}(1 + y - x) \quad (4.34)$$

is both improving and IM. The following remark follows from Proposition 12 (and Proposition 9).

Remark 7 *If every individual in the population use the same IM rule, then $p'(A^*) \geq p(A^*)$ in every environment, with strict inequality when $p(A^*) \in (0, 1)$.*

G. Second-Order Monotonicity

A natural extension of the analysis provided above relates to the concept of second-order stochastic dominance. When payoffs are monetary, very often decision makers are concerned about risk. Usually this concern leads them to refrain from playing actions providing higher expected payoffs when they are associated with a higher

risk. Are there specific features that one can impose on imitation rules in order to lead the population to rational choice consistent with risk aversion? In order to introduce such concerns, a similar analysis to the one provided for first-order stochastic dominance and FOM rules can be developed for second-order stochastic dominance. This analysis reveals the conditions that need to be imposed on rules so that they lead the population to make safer choices. Formally, we can characterize a family of rules such that, in environments where all the actions have the same expected payoffs, the expected fraction of the population playing in the next period an action that second-order stochastically dominates all the other actions in the set is higher than that fraction in the current period. We call such rules *second-order monotone* (SOM) and we characterize them. The characterization of SOM rules is analogous to the characterization of FOM rules, except that the net-switching functions are concave instead of increasing in the payoff of the action that receives the probability. It follows that an interesting subset of FOM rules is the one with rules that are both FOM and SOM at the same time, i.e., those with increasing, concave net-switching functions. These rules lead to rational choice in the sense of first-order monotonicity and, at the same time, are consistent with risk-averse decision making.

In order to isolate risk attitudes, in this section we consider properties of rules in environments where the expected payoff of the distributions of different actions is the same. This is the only class of environments we consider in this section. In what follows, second-order stochastic dominance, abbreviated by $F_a \text{ sosd } F_{a'}$, means that $\int_0^x F_{a'}(t) - F_a(t) dt \geq 0$ for all $x \in [0, 1]$. We start by defining the set of actions that would be preferred for risk-averse decision makers, i.e., the set of actions whose payoff distribution second-order stochastically dominates the distributions of the other actions in the set. Let $A_* := \{a \in A : F_a \text{ sosd } F_{a'} \forall a' \in A\}$.

Definition 19 A rule L is said to be second-order monotone (SOM) if $p'(A_*) \geq p(A_*)$ in every environment where all the actions have the same expected payoff.

Since the structure of the definition of SOM rules is analogous to that of FOM rules, they can be characterized in an analogous way. This characterization is provided in Lemma 10 and Proposition 13¹¹.

Lemma 10 A rule L is SOM if and only if it satisfies the following conditions:

- (i) L is imitative,
- (ii) $F_{a'} \text{ sosd } F_a \Rightarrow L_{aa'}^{a'} - L_{a'a}^a \geq 0, \forall a, a' \in A$, in every environment.

Proof. Sufficiency is proved analogously to the proof of sufficiency in Lemma 1.

Necessity of (i):

Consider the SOM rule L . Let $x, y \in [0, 1]$; $a, a', a'' \in A, a'' \notin \{a, a'\}$. Suppose $L(a, x, a', y)_{a''} > 0$. Consider an environment where $F_a = F_{a'}$, $\mu_a(x) = \mu_a(y) = \mu_a(1/2) = 1/3$, and for all $a''' \in A \setminus \{a, a'\}$, $\mu_{a'''}(x) = \mu_{a'''}(y) = 1/3$ and $\mu_{a'''}(1/4) = \mu_{a'''}(3/4) = 1/6$. It follows that $A_* = \{a, a'\}$. Assume that $p(A_*) = 1$. Suppose that there are $c, d \in W$ such that $\Pr(cd) > 0$, $s(c) = a$, and $s(d) = a'$, then $p'(A_*) < 1$. Thus, L has to be imitative.

Necessity of (ii):

Suppose that for some $a, a' \in A$, $F_{a'} \text{ sosd } F_a$, but $L_{aa'}^{a'} - L_{a'a}^a < 0$. Suppose $a' \in A^*$. Consider $\varepsilon \in (0, 1)$. If $\mu_a(1) = 1$ for all $a \in A$, then consider the modified environment \hat{F} such that $\hat{\mu}_a(1) = 1 - \varepsilon$ and $\hat{\mu}_a(y) = \varepsilon$ for some $y \in [0, 1]$, for all $a \in A$. If $\mu_a(0) = 1$ for all $a \in A$, then consider the modified environment \hat{F} such that $\hat{\mu}_a(0) = 1 - \varepsilon$ and $\hat{\mu}_a(y) = \varepsilon$ for some $y \in (0, 1]$, for all $a \in A$. Otherwise just let \hat{F}

¹¹As in EUT, the analysis of SOM rules is easier to carry on when properties are stated in terms of weak inequalities.

$= F$. Denote by π the expected value of the distributions in the environment \widehat{F} . Now consider the new modified environment \widetilde{F} such that for any interval $I \subseteq [0, 1] \setminus \{\pi\}$, $\widetilde{\mu}_{a'}(I) = (1 - \varepsilon)\widehat{\mu}_{a'}(I)$, $\widetilde{\mu}_{a'}(\pi) = \widehat{\mu}_{a'}(\pi) + \varepsilon\widehat{\mu}_{a'}([0, 1] \setminus \{\pi\})$; for any interval $I \subseteq (0, 1)$, $\widetilde{\mu}_{a'''}(I) = (1 - \varepsilon)\widehat{\mu}_{a'''}(I)$, $\widetilde{\mu}_{a'''}(1) = (1 - \varepsilon)\widehat{\mu}_{a'''}(1) + \varepsilon\pi$, and $\widetilde{\mu}_{a'''}(0) = (1 - \varepsilon)\widehat{\mu}_{a'''}(0) + \varepsilon(1 - \pi)$ for all $a''' \in A \setminus \{a'\}$. In this new environment $\widetilde{A}_* = \{a'\}$. Now, conclude as in Lemma 7. \square

From Lemma 10, it is clear that SOM rules are impartial.

Proposition 13 *L is SOM if and only if it satisfies the following conditions:*

(i) *L is imitative,*

(ii) *for all $a, a' \in A$, the net-switching function $g_{aa'}(x, y)$ satisfies:*

$$(ii.1) \quad g_{aa'}(x, y) = -g_{aa'}(y, x), \quad \forall x, y \in [0, 1]$$

$$(ii.2) \quad g_{aa'}(x, y) \text{ is concave in } y, \text{ for all } x \in [0, 1].$$

Proof. Sufficiency and the necessity of the condition (ii.1) can be proven in the same way as in Proposition 9. Now we prove that $g_{aa'}(x, y)$ is concave with respect to y for SOM rules. Suppose that for some $x, y', y, \lambda \in [0, 1]$, and $y'' \equiv \lambda y + (1 - \lambda)y'$ we have that

$$g_{aa'}(x, y'') < \lambda g_{aa'}(x, y) + (1 - \lambda)g_{aa'}(x, y'). \quad (4.35)$$

Consider an environment where $\mu_a(x) = \mu_{a'}(x) = 1 - \varepsilon$, $\mu_a(y) = \lambda\varepsilon$, $\mu_a(y') = \varepsilon(1 - \lambda)$, $\mu_{a'}(y'') = \varepsilon$, and notice that $F_{a'}$ sosd F_a . Lemma 10 and condition (ii.1) imply

$$\begin{aligned} 0 &\leq L_{aa'}^{a'} - L_{a'a}^a \\ &= (1 - \varepsilon)^2 g_{aa'}(x, x) + (1 - \varepsilon)\varepsilon g_{aa'}(x, y'') + (1 - \varepsilon)\lambda\varepsilon g_{aa'}(y, x) \\ &\quad + (1 - \varepsilon)\varepsilon(1 - \lambda)g_{aa'}(y', x) + \lambda\varepsilon^2 g_{aa'}(y, y'') + (1 - \lambda)\varepsilon^2 g_{aa'}(y', y'') \quad (4.36) \\ &= \varepsilon[\lambda\varepsilon g_{aa'}(y, y'') + (1 - \lambda)\varepsilon g_{aa'}(y', y'') \\ &\quad + (1 - \varepsilon)\{g_{aa'}(x, y'') - \lambda g_{aa'}(x, y) - (1 - \lambda)g_{aa'}(x, y')\}]. \end{aligned}$$

By our hypothesis, the term inside the brackets $\{.\}$ is negative. Thus, for small enough ε , the term inside the brackets $[.]$ is negative, causing a contradiction. \square

The following corollary completes the analogy between the characterizations of FOM and SOM rules and it follows directly from condition (ii) of Proposition 13.

Corollary 5 *If L is SOM then $g_{aa'}(x, y)$ is convex in x , for all $y \in [0, 1]$.*

H. Discussion

In this section we consider possible directions for future research. As explained in the introduction, our motivation to study FOM rules was identifying a set of learning rules that lead individuals to choose actions in a way that is consistent with the decisions of fully-informed rational agents. However, one can think about a motivation for FOM rules based on arguments related to evolutionary theory. For example, Schlag's [33] improving rules are motivated by their performance under selection pressure. If survival of imitation rules depends on their average payoff, then a successful rule must be able to lead decision towards expected payoff maximising actions. An interesting direction for further research is finding the conditions under which FOM rules may prevail under selection pressure. More specifically, it would be interesting to find the specific ways (if any) in which survival must be determined in order to allow FOM rules to be dominant. The analysis in Robson [27] is also related.

A restrictive feature in our analysis is the assumption that the payoff-distributions associated to the different actions are the same for all individuals in the population. In many settings one would like to allow for different payoff distributions for agents with different characteristics. In Ellison and Fudenberg [14] heterogeneity of the population makes it harder for a social learning process to lead the population to play the action that provides the highest expected payoff. Future research could extend the

analysis to allow for this possibility and try to identify rules that lead to (expected) better decisions in the future in heterogeneous environments. Future research could also identify and characterize properties based on notions analogous to the ones studied in this essay, but suitable to the information contexts of different models, for example when two or more other individuals are sampled.

Our concept of FOM rules serves to provide a foundation for a number of imitation rules in the literature, including very simple rules such as Imitate if Better and relatively more sophisticated rules such as the Proportional Imitation Rule. We show that among all the FOM rules, none of them outperforms all the others in every environment. Therefore we do not single out a best rule. This non-uniqueness parallels non-uniqueness in rational choice. In our view, a next step in order to bound the set of imitation rules should be based on experimental evidence. Future research could test if experimental subjects use imitation rules that are FOM and try to find out what are the specific shapes of the rules they use. The hypothesis of risk-averse imitation may be of interest as well.

CHAPTER V

SUMMARY

In the three essays of this dissertation we have shown how the concept of risk aversion can be extended to the analysis of decisions beyond the scope of Expected Utility Theory. Decision makers do not need to know the payoff distributions of the different alternatives to display behavior consistent with a suitable notion of risk aversion.

In the model we study, decision makers respond, in expected terms, to risky distributions in a way that leads them to be less likely to make choices that are riskier. This is a consequence of the way their decisions respond to payoffs. If the probability of choosing an action responds in a concave way to the obtained payoff with that action, then it is more likely that in the future these individuals will play safer actions and will move away from risky actions. We have characterized the learning rules satisfying these properties in terms of both second-order stochastic dominance and the traditional concept of variance. This analysis was provided in Chapter II and Chapter III respectively. When the analysis of risk is carried on in terms of stochastic dominance, the response of the probability to obtained payoffs must be a concave function for the action that has been played. When the analysis of risk is conducted in terms of variance, such transformation must be quadratic. This is consistent with the analysis of Bernoulli utility functions in Expected Utility analysis where these functions must be concave to represent risk aversion and quadratic to represent mean-variance preferences. Chapter IV of this dissertation extends the analysis to study these issues when learning occurs in social contexts. Here, individuals learn from their own choices and payoffs and from the choices and payoffs of other individuals in the population. In order to display risk aversion, imitation rules must respond in a concave way to the payoffs of the actions that they may imitate.

While our analysis provides a significant step towards understanding how risk aversion may be thought of in contexts that do not fit the Expected Utility paradigm, we have not, as yet, illustrated how these ideas can be incorporated to the analysis of decision in problems of general interest in economics; or how these ideas may provide further insights in contexts that fit well the framework of our analysis. We leave the pursue of this agenda for future research.

REFERENCES

- [1] J. Apesteguia, S. Huck, and J. Oechssler, Imitation–Theory and Experimental Evidence, *J. Econ. Theory*, forthcoming.
- [2] K.J. Arrow, *Essays in the Theory of Risk-Baring*, Markham, Chicago, IL, 1974.
- [3] T. Börgers, On the Relevance of Learning and Evolution to Economic Theory, *Econ. J.* 106 (1996), 1375-1385.
- [4] T. Börgers, A. Morales, R. Sarin, Expedient and Monotone Learning Rules, *Econometrica* 72 (2004), 383-405.
- [5] T. Börgers, R. Sarin, Learning Through Reinforcement and Replicator Dynamics, *J. Econ. Theory* 77 (1997), 1-14.
- [6] A. Burgos, Learning to Deal with Risk: What Does Reinforcement Learning Tell Us About Risk Attitudes?, *Econ. Bull.* 4 (2002), 1-13.
- [7] C. Camerer, T.H. Ho, Experience-Weighted Attraction Learning in Normal-Form Games, *Econometrica* 67 (1999), 827-874.
- [8] J. Cross, A Stochastic Learning Model of Economic Behavior, *Quart. J. Econ.* 87 (1973), 239-266.
- [9] R. Cubitt, R. Sugden, The Selection of Preferences Through Imitation, *Rev. Econ. Stud.* 65 (1998), 761-771.
- [10] E. Dekel, S. Scotchmer, On the Evolution of Attitudes Towards Risk in Winner-Take-All games, *J. Econ. Theory* 87 (1999), 125-143.

- [11] S. Della Vigna, M. Li Calzi, Learning to Make Risk-Neutral Choices in a Symmetric World, *Math. Soc. Sci.* 41 (1999), 19-37.
- [12] J. Denrell, Adaptive Learning and Risk Taking, *Psychol. Rev.* 114 (2007), 177-187.
- [13] J. Denrell, J.G. March, Adaptation as Information Restriction: The Hot Stove Effect, *Organ. Sci.* 12 (2001), 523-538.
- [14] G. Ellison, D. Fudenberg Rules of Thumb for Social Learning, *J. Polit. Economy* 101 (1993), 612-643.
- [15] G. Ellison, D. Fudenberg, Word of Mouth Communication and Social Learning, *Quart. J. Econ.* 110 (1995), 93-125.
- [16] I. Erev, Y. Bereby-Meyer, A. Roth, The Effect of Adding a Constant to All Payoffs: Experimental Investigation and Implications for Reinforcement Learning, *J. Econ. Behav. Organ.* 39 (1999), 111-128.
- [17] D. Fudenberg, D. Levine, *The Theory of Learning in Games*, MIT Press, Cambridge, MA, 1998.
- [18] F. Gul, W. Pesendorfer, Random Expected Utility, *Econometrica* 74 (2006), 121-146.
- [19] G. Hanoch, H. Levy, The Efficiency Analysis of Choices Involving Risk, *Rev. Econ. Stud.* 36 (1969), 335-346.
- [20] E. Haruvy, I. Erev, D. Sonsino, The Medium Prizes Paradox: Evidence from a Simulated Casino, *J. Risk Uncertainty* 22 (2001), 251-261.

- [21] E. Karni, D. Schmeidler, Self-Preservation as a Foundation of Rational Behavior Under Risk, *J. Econ. Behav. Organ.* 7 (1986), 71-81.
- [22] J.G. March, Learning to Be Risk Averse, *Psychol. Rev.* 103 (1996), 309-319.
- [23] H. Markowitz, Portfolio Selection, *J. Financ.* 7 (1952), 77-91.
- [24] J. Meyer, Two Moment Decision Models and Expected Utility Maximization, *Amer. Econ. Rev.* 77 (1987), 421-430.
- [25] A. Morales, Absolutely Expedient Imitative Behavior, *Int. J. Game Theory* 31 (2002), 475-492.
- [26] J.W. Pratt, Risk Aversion in the Small and in the Large, *Econometrica* 32 (1964), 122-136.
- [27] A.J. Robson, A Biological Basis for Expected and Non-Expected Utility Preferences, *J. Econ. Theory* 68 (1996), 397-424.
- [28] A.J. Robson, The Evolution of Attitudes Towards Risk: Lottery Tickets and Relative Wealth, *Game. Econ. Behav.* 14 (1996), 190-207.
- [29] A.E. Roth, I. Erev, Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term, *Game. Econ. Behav.* 8 (1995), 164-212.
- [30] M. Rothschild, A Two-Armed Bandit Theory of Market Pricing, *J. Econ. Theory* 9 (1974), 185-202.
- [31] M. Rothschild, J.E. Stiglitz, Increasing Risk: I. A Definition, *J. Econ. Theory* 2 (1970), 225-243.

- [32] A. Rustichini, Optimal Properties of Stimulus-Response Learning Models, *Game. Econ. Behav.* 29 (1999), 244-273.
- [33] K. Schlag, Why Imitate, and if so, How? A Bounded Rational Approach to Multi-Armed Bandits, *J. Econ. Theory* 78 (1998), 130-156.
- [34] K. Schlag, Which One Should I Imitate,? *J. Math. Econ.* 31 (1999), 493-522.
- [35] H.A. Simon, A Comparison of Game Theory and Learning Theory, *Psychometrika* 21 (1956), 267-272.
- [36] J. Tobin, Liquidity Preference as Behavior Toward Risk, *Rev. Econ. Stud.* 25 (1958), 65-86.
- [37] F. Vega-Redondo, The Evolution of Walrasian Behavior, *Econometrica* 65 (1997), 375-384.

VITA

Carlos Oyarzun was born in Concepcion, Chile, in 1974. He received his Bachelor of Arts degree in Economics from the Universidad de Concepcion in 1995 and Master of Arts degree in Economics from the Universidad de Chile in 1999. He entered the graduate program in economics at Texas A&M University in August 2002, and received his Doctor of Philosophy degree in August 2007. Under the supervision of Dr. Rajiv Sarin, his research interests include Behavioral Decision Theory, Game Theory, and Learning Theory. Dr. Oyarzun may be reached at Departamento de Fundamentos del Analisis Economico, Universidad de Alicante, Campus de San Vicente, 03080 Alicante, Spain. His email address is oyarzun@merlin.fae.ua.es.